

Note di probabilità e statistica  
per il corso di Laboratorio 2

Anna Martin  
Dipartimento di Fisica, Università degli Studi di Trieste

a.a. 2020/21, v011222, rev capitolo 1

\*

Queste note sintetiche sono una raccolta di argomenti di teoria della probabilità e di statistica trattati durante il corso di Laboratorio 2, ad uso degli studenti che seguono il corso. Non hanno alcuna pretese di completezza degli argomenti trattati, come ovvio, visto il tempo a disposizione.

Non sono inclusi alcuni argomenti di teoria della probabilità trattati nel corso di Laboratorio 1 e che vengono dati per scontati (ad es. concetti di fenomeno statistico e di variabili casuali, definizioni di probabilità, teorema di Bayes, introduzione alle distribuzioni binomiale e di Poisson, e alle funzioni di distribuzione uniforme e di Gauss).

Pur essendo parte integrante del programma e fondamentali per la comprensione degli argomenti trattati, in queste note non ci sono alcune dimostrazioni, esempi e applicazioni fatti a lezione, e la parte pratica del corso, che cambia di anno in anno. Per questi punti, che fanno parte integrale del programma, si rimanda alla pagina web e agli appunti presi durante il corso.

Altro materiale di consultazione è costituito dai molti testi esistenti <sup>1</sup>

Infine, un ringraziamento speciale a Francesca Cucchiario e Simone Schiavon, che, nell'a.a. 2020/21, hanno rivisto e integrato la prima parte di queste note. E grazie a chi, in futuro, leggendo queste note, mi segnalerà errori o imprecisioni ....

---

<sup>1</sup>Per citarne alcuni:

- F. Bradamante, "Note di analisi statistica dei dati", alle quali queste note sono ispirate
- G. Cowan, *Statistical Data Analysis*, Clarendon Press, Oxford
- A. Rotondi, P. Pedroni, A. Pievatolo, "Probabilità Statistica e Simulazione", Springer
- review di statistica da R.L. Workman et al. (Particle Data Group), *Prog. Theor. Exp. Phys.* 2022, 083C01 (2022) <http://pdg.lbl.gov/2015/reviews/rpp2014-rev-statistics.pdf>

## Introduzione

La statistica <sup>2</sup> è una disciplina che ha come fine lo studio quantitativo e qualitativo di un particolare fenomeno collettivo in condizioni di incertezza o non determinismo. È uno strumento del metodo scientifico, che si avvale della matematica per studiare i modi in cui un fenomeno collettivo può essere sintetizzato e compreso e ciò avviene attraverso la raccolta e l'analisi delle informazioni relative al fenomeno studiato. Viene comunemente suddivisa in due branche principali:

- descrittiva:  
sintetizza i dati attraverso grafici (diagrammi a barre, a torta, istogrammi, ecc.) e indici (indicatori di posizione, di dispersione, di forma) che descrivono gli aspetti salienti dei dati osservati;
- inferenziale:  
ha come obiettivo quello di stabilire delle caratteristiche dei dati e dei comportamenti delle misure (variabili statistiche) con una possibilità di errore predeterminata. Possono riguardare la natura (la legge probabilistica) del fenomeno che si osserva, la cui conoscenza permetterà poi di fare una previsione. La statistica inferenziale è fortemente legata alla teoria della probabilità, in quanto descrivere in termini probabilistici o statistici un fenomeno aleatorio, vuol dire descriverlo in termini di distribuzione di probabilità e dei suoi parametri.

La statistica inferenziale si suddivide poi in altri capitoli, di cui tra i più importanti ci sono

- la teoria della stima, puntuale e intervallare, ("stima dei parametri"):  
consiste nell'estrarre da un campione casuale rappresentativo informazioni sull'intera popolazione, come i valori dei parametri della distribuzione di una variabile casuale oppure dei parametri che compaiono in una relazione tra variabili casuali;
- la verifica delle ipotesi ("test d'ipotesi"):  
consiste nell'estrarre informazioni quantitative sulla validità di un'ipotesi statistica (funzione di distribuzione di variabili casuali, relazione tra variabili casuali) a partire da un campione rappresentativo.

La stima dei parametri e il test d'ipotesi sono il punto di arrivo del corso. Naturalmente verranno trattati solo alcuni aspetti di base di questi importanti capitoli, limitandosi alla inferenza statistica "classica". Da quanto detto, al di là delle terminologia usata, dovrebbe comunque essere chiaro perchè tutta la prima parte è relativa alla teoria della probabilità, che costituisce la base di partenza.

---

<sup>2</sup>definizioni tratte da Wikipedia

# 1 Teoria della probabilità

## 1.1 Fenomeni statistici caratterizzati da una variabile casuale

In queste note,

Una variabile casuale è una grandezza fisica di cui a priori non si può conoscere il valore (però si può ad esempio conoscerne la funzione di distribuzione). Tale incertezza è dovuta a due motivi principali:

- la grandezza fisica ha caratteristiche intrinsecamente statistiche cioè non possiede un determinato "valore vero";
- la grandezza fisica possiede un "valore vero" ma l'operazione di misura introduce errori accidentali che determinano valori diversi per ogni misura.

### 1.1.1 Variabili casuali discrete

La variabile casuale  $x$ <sup>3</sup> è discreta se può assumere solo valori appartenenti a un insieme numerabile  $\Omega_x = \{x_1, \dots, x_n\}$  con  $n$  che può essere anche infinito. In particolare, la probabilità che  $X$  assuma il valore  $x_k$  è data da:

$$P_k = P(x = x_k) \quad k = 1, \dots, n.$$

L'insieme dei valori  $P_k$  costituisce la distribuzione di probabilità  $\{P_k\}$ . Per definizione di probabilità, deve essere:

1. non negativa e limitata:  $0 \leq P_k \leq 1$
2. normalizzata:  $\sum_{k=1}^n P_k = 1$

Valore di aspettazione e varianza sono definiti, rispettivamente, come

$$E[x] = \mu_x = \sum_{k=1}^n x_k P_k, \quad var(x) = \sigma_x^2 = E[(x - \mu_x)^2] = \sum_{k=1}^n (x_k - \mu_x)^2 P_k,$$

ed è anche  $var(x) = E[x^2] - E[x]^2$ .

Distribuzioni di probabilità di variabili casuali discrete sono, ad esempio, la distribuzione binomiale e la distribuzione di Poisson.

---

<sup>3</sup>In alcuni testi la variabile casuale viene indicata utilizzando una lettera maiuscola ( $X$ ) mentre l'uso della minuscola ( $x$ ) è riservato per gli specifici valori che la variabile casuale può assumere. Qui, in genere, viene usata la lettera minuscola in entrambi i casi, assumendo che il suo significato sia chiaro dal contesto.

### 1.1.2 Variabili casuali continue

Una variabile casuale  $x$  è continua se assume valori appartenenti a un intervallo di numeri reali  $\Omega_x = (a, b)$  con eventualmente  $a = -\infty$  e  $b = +\infty$ . Dal momento che  $x$  può assumere tutti i valori dell'intervallo, non ha senso assegnare un valore di probabilità a ciascun valore cioè  $P(x = x_k) = 0$ . Invece la probabilità che  $x$  cada in un certo intervallo  $[x_1, x_2]$  è data da:

$$P(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} f(x)dx$$

dove  $f(x)$  prende il nome di funzione di distribuzione (o densità di probabilità) di  $X$ .

Per definizione di probabilità la funzione di distribuzione deve soddisfare alcune importanti proprietà. In particolare deve essere

1. limitata e definita positiva in  $\Omega_x$ :  $0 \leq f(x) \leq 1 \quad \forall x \in \Omega_x$
2. normalizzata:  $\int_{\Omega_x} f(x)dx = 1$ .

In analogia al caso delle variabili casuali discrete, valore di aspettazione e varianza sono definiti come:

$$E[x] = \mu_x = \int_{\Omega_x} x \cdot f(x)dx, \quad var(x) = \sigma_x^2 = E[(x - \mu_x)^2] = \int_{\Omega_x} (x - \mu_x)^2 \cdot f(x)dx,$$

e, come si può facilmente verificare, è ancora  $var(x) = E[x^2] - E[x]^2$ .

La deviazione standard (errore statistico)  $\sigma_x$  è poi la radice quadrata della varianza.

Nel seguito, molti argomenti verranno trattati riferendosi a variabili casuali continue, ma la loro formulazione nel caso di variabili casuali discrete dovrebbe essere in enere semplice.

### 1.1.3 Varianza e disuguaglianza di Chebyshev

Il fatto che come indice di dispersione si introduca la varianza è dovuto alla validità di una importante disuguaglianza, che permette di porre dei limiti alla probabilità che la variabile casuale assuma valori in determinati intervalli (o insiemi di valori) qualunque sia la sua funzione di distribuzione (o distribuzione di probabilità, purchè la varianza sia finita. Si tratta della *disuguaglianza di Chebyshev* che afferma che :

$$P(|x - \mu| \geq \lambda \sigma_x) \leq \frac{1}{\lambda^2}$$

indipendentemente dalla funzione di distribuzione purchè la varianza sia finita.

\* Dimostrazione per una variabile casuale continua:

se  $f(x)$  è la funzione di distribuzione di  $x$ , si introduce la funzione  $h(x) \geq 0$  sull'intervallo di definizione di  $x$ . Fissato un valore  $k \geq 0$  compreso tra il valore minimo e il valore massimo di  $h(x)$  in  $\Omega_x$ , e detto  $R$  il sottoinsieme di  $\Omega_x$  in cui  $h(x) \geq k$ , si ha:

$$E[h(x)] = \int_{\Omega_x} h(x) \cdot f(x) dx \geq \int_R h(x) \cdot f(x) dx \geq k \int_R f(x) dx = kP(h(x) \geq k)$$

$$\implies P(h(x) \geq k) \leq \frac{E[h(x)]}{k}$$

ponendo  $k = (\lambda\sigma)^2$ ,  $h(x) = (x - \mu)^2$  si ottiene la tesi.

Si può facilmente vedere che la diseuguaglianza è verificata ad esempio per la funzione di distribuzione di Gauss. La sua importanza sta nel fatto che è di validità generale, e la useremo nel seguito.

#### 1.1.4 Momenti e funzioni generatrici dei momenti

I momenti sono i valori di aspettazione (quindi quantità numeriche) di particolari funzioni della variabile casuale  $X$ , che caratterizzano le funzioni di distribuzione.

I *momenti algebrici* di ordine  $k$  sono definiti come:

$$\mu_k^* = E[x^k] = \int_{\Omega_x} x^k \cdot f(x) dx$$

dove  $f(x)$  è la funzione di distribuzione di  $x$ .

Si ha quindi:

$$\mu_0^* = \int_{\Omega_x} 1 \cdot f(x) dx = E[1] = 1 \quad \rightarrow \text{condizione di normalizzazione di } f(x)$$

$$\mu_1^* = \int_{\Omega_x} x \cdot f(x) dx = E[x] = \mu_x \quad \rightarrow \text{indice di posizione}$$

$$\mu_2^* = \int_{\Omega_x} x^2 \cdot f(x) dx = E[x^2] \quad \rightarrow \text{valore di aspettazione di } x^2$$

....

Il significato di  $\mu_2^*$  e dei momenti algebrici di ordine superiore non è immediato.

I *momenti centrali* di ordine  $k$  sono riferiti al valore di aspettazione di  $x$ , e sono definiti come:

$$\mu_k = E[(x - \mu_x)^k] = \int_{\Omega_x} (x - \mu_x)^k \cdot f(x) dx$$

Si ha quindi:

$$\mu_0 = \int_{\Omega_x} 1 \cdot f(x) dx = E[1] = 1 \quad \rightarrow \text{condizione di normalizzazione di } f(x)$$

$$\mu_1 = \int_{\Omega_x} (x - \mu_x) \cdot f(x) dx = E[x - \mu_x] = 0 \quad \rightarrow \text{valore di aspettazione di } x - \mu_x$$

$\mu_2 = \int_{\Omega_x} (x - \mu_x)^2 \cdot f(x) dx = E[(X - \mu_x)^2] = \sigma_x^2 \rightarrow$  *indice di dispersione*; già commentato

$\mu_3 = \int_{\Omega_x} (x - \mu_x)^3 \cdot f(x) dx = E[(X - \mu_x)^3] \rightarrow$  *indice di asimmetria*; è un indice di forma, ed è chiaramente uguale a zero se  $f(x)$  è simmetrica rispetto a  $\mu_x$

$\mu_4 = \int_{\Omega_x} (x - \mu_x)^4 \cdot f(x) dx = E[(X - \mu_x)^4] \rightarrow$  *indice di curtosi*; altro indice di forma, dà informazioni su quanto è "piatta" la funzione di distribuzione (o "alte" le code)

...

Per i momenti centrali di ordine superiore a 2, si introducono i *coefficienti*.

In particolare, si definisce *coefficiente di asimmetria* (skewness) la quantità adimensionale:

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma_x^3}$$

e *coefficiente di curtosi* (kurtosis) la quantità adimensionale:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$$

dove viene sottratto 3 per avere  $\gamma_2 = 0$  per la funzione di distribuzione di Gauss (vedi dopo).

Per capire meglio il significato di questi coefficienti, è utile confrontare i loro valori per la funzione di distribuzione di Gauss, uniforme, ed esponenziale (vedi sezioni dedicate a queste funzioni).

I momenti possono tutti essere ottenuti dalle derivate di particolari funzioni, le *funzioni generatrici* dei momenti algebrici e centrali, definite rispettivamente come:

$$M_x^*(t) = E[e^{tx}] = \int_{\Omega_x} e^{tx} f(x) dx$$

$$M_x(t) = E[e^{t(x-\mu_x)}] = \int_{\Omega_x} e^{t(x-\mu_x)} f(x) dx$$

dove  $t$  è una variabile reale. Le due funzioni sono legate dalla relazione:

$$M_x(t) = e^{-t\mu_x} M_x^*(t).$$

La connessione tra momenti e funzioni generatrici è evidente quando si sviluppano le funzioni esponenziali in serie di potenze:

$$M_x^*(t) = E[e^{tx}] = \int_{\Omega_x} e^{tx} f(x) dx = \sum_{k=0}^{\infty} \frac{t^k}{k!} \int_{\Omega_x} x^k \cdot f(x) dx = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mu_k^*$$

e analoga per  $M_x(t)$ . Vale quindi:

$$\mu_k = \left[ \frac{\partial^k M_x}{\partial t^k} \right]_{t=0} \quad \mu_k^* = \left[ \frac{\partial^k M_x^*}{\partial t^k} \right]_{t=0}$$

Dalle funzioni generatrici si possono quindi ottenere tutti i momenti, cioè le funzioni generatrici contengono le informazioni su tutti i momenti.

Esistono casi in cui le funzioni generatrici non sono calcolabili (distribuzione di Cauchy). In questi casi però si possono calcolare le *funzioni caratteristiche*:

$$\phi_x^*(t) = E[e^{itx}] \implies i^k \mu_k^* = \left[ \frac{\partial^k \phi_x^*}{\partial t^k} \right]_{t=0}$$

L'importanza delle funzioni generatrici sta anche nel fatto che, contenendo tutti i momenti, caratterizzano completamente le funzioni di distribuzioni.

Si può infatti dimostrare che, in condizioni molto generali, l'insieme di tutti i momenti caratterizza completamente la funzione di distribuzione.

In particolare, se due funzioni di distribuzione  $f_1(x)$  e  $f_2(x)$  possono essere sviluppate in serie di potenze e se hanno lo stesso insieme di momenti, allora queste coincidono:  $f_1(x) = f_2(x)$ .

\* Dimostrazione:

Date  $f_1(x) = \sum_{i=0}^{\infty} a_i x^i$  e  $f_2(x) = \sum_{j=0}^{\infty} b_j x^j$ , si ha

$f_1(x) - f_2(x) = \sum_{k=0}^{\infty} (a_k - b_k) x^k = \sum_{k=0}^{\infty} c_k x^k$  da cui

$[f_1(x) - f_2(x)]^2 = \sum_{k=0}^{\infty} c_k x^k [f_1(x) - f_2(x)]$  e, integrando su  $\Omega_x$ :

$$\begin{aligned} \int_{\Omega_x} [f_1(x) - f_2(x)]^2 dx &= \sum_{k=0}^{\infty} c_k \int_{\Omega_x} x^k [f_1(x) - f_2(x)] dx \\ &= \sum_{k=0}^{\infty} c_k \left( \int_{\Omega_x} x^k f_1(x) dx - \int_{\Omega_x} x^k f_2(x) dx \right) = \sum_{k=0}^{\infty} c_k [\mu_k^{1*} - \mu_k^{2*}] = 0 \end{aligned}$$

$\implies f_1(x) = f_2(x)$  su  $\Omega_x$ .

Quindi se due funzioni di distribuzione hanno le stesse funzioni generatrici dei momenti allora coincidono. Tale proprietà verrà utilizzata nel seguito, quando una funzione generatrice dei momenti per una variabile casuale con funzione di distribuzione non nota risulterà essere la funzione generatrice nota di una certa funzione di distribuzione. Naturalmente questo implica l'aver già calcolato le funzioni generatrici ...

## 1.2 Alcune distribuzioni di probabilità e funzioni di distribuzione

Questa sezione è dedicata ad alcune distribuzioni di probabilità e funzioni di distribuzione particolarmente diffuse. Per quelle già note da corsi precedenti (distribuzione binomiale e di Poisson, funzioni di distribuzione uniforme e di Gauss) sono omesse derivazioni e considerazioni generali sulle loro caratteristiche, considerate note.

### 1.2.1 Funzione di distribuzione di Gauss

La funzione di distribuzione di Gauss (o normale):

$$f(x) = N(\mu; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

è fondamentale, in quanto molto diffusa in fisica e non solo. La variabile casuale  $x$  può assumere valori su tutto l'asse reale e si può dimostrare che

- la funzione di distribuzione è normalizzata:  $\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$
- il valore di aspettazione è  $E[x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu$
- la varianza è  $var(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x^2 \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx - \mu^2 = \sigma^2$

Nel caso in cui  $\mu = 0$  e  $\sigma^2 = 1$ , allora si parla di *distribuzione normale standard*:

$$f(x) = N(0; 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Esempi di funzione di distribuzione di Gauss per diversi valori dei parametri  $\mu$  e  $\sigma^2$  sono riportati in fig. 1.

I valori della funzione di ripartizione come pure gli integrali della funzione di distribuzione  $N(0,1)$  in specifici intervalli centrati in zero sono tabulati, e sono disponibili anche "calcolatori" online.

Come si può calcolare facilmente, le funzioni generatrici dei momenti sono:

$$M_x(t) = E \left[ e^{t(x-\mu)} \right] = e^{\sigma^2 t^2 / 2}, \quad M_x^*(t) = e^{t\mu} M_x(t) = e^{\mu t + \sigma^2 t^2 / 2}$$

\* Per calcolare  $M_x(t) = E \left[ e^{t(x-\mu)} \right]$  basta riscrivere l'esponente:  
 $t(x-\mu) - (x-\mu)^2 / 2\sigma^2 = -[(x-\mu) - \sigma^2 t]^2 / 2\sigma^2 + \sigma^2 t^2 / 2$   
da cui

$$M_x(t) = e^{\sigma^2 t^2 / 2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}} dx = e^{\sigma^2 t^2 / 2}$$

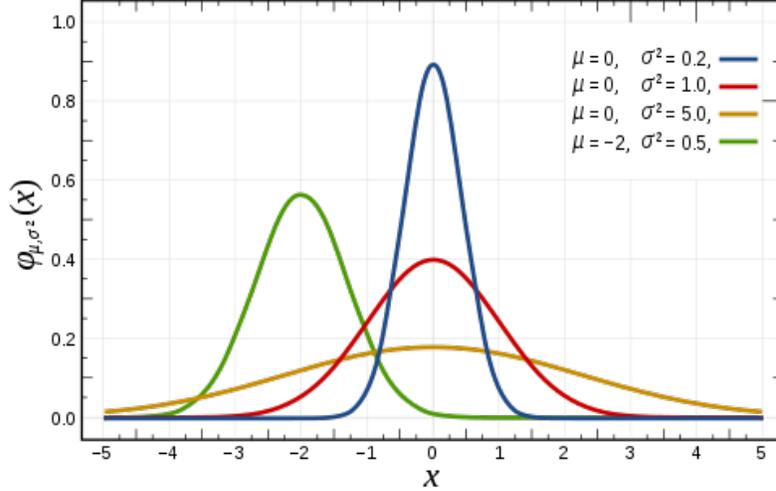


Figura 1: Funzioni di distribuzione di Gauss per diversi valori di  $\mu$  e  $\sigma^2$ .

Per la distribuzione normale standard  $N(0; 1)$  si ha quindi:

$$M_x(t) = e^{t^2/2} = M_x^*(t)$$

Per la generica funzione di distribuzione normale, calcolando le derivate di  $M_k^*(x)$  in  $t = 0$  si ottengono i momenti algebrici:

$$\mu_0^* = 1, \quad \mu_1^* = E[x] = \mu \Delta$$

come previsto.

Calcolando le derivate di  $M_k(x)$  in  $t = 0$  si ottengono i momenti centrali (i calcoli non presentano particolari difficoltà):

$$\mu_0 = 1, \quad \mu_1 = 0, \quad \mu_2 = var(x) = \sigma^2$$

anche in questo caso, come previsto. Per i momenti centrali di ordine 3 e 4, e per i coefficienti di asimmetria e di curtosi si ottiene:

$$\mu_3 = 0 \implies \gamma_1 = 0, \quad \mu_4 = 3\sigma^4 \implies \gamma_2 = 0.$$

Questi sono i valori di riferimento, in quanto le caratteristiche delle altre funzioni di distribuzione vengono in genere confrontate con quelle della funzione di distribuzione di Gauss.

### 1.2.2 Funzione di distribuzione uniforme

Una variabile casuale continua ha distribuzione uniforme in  $(a, b)$  se:

$$f(x) = \begin{cases} c & \text{se } x \in (a, b) \\ 0 & \text{se } x \notin (a, b) \end{cases}$$

Si trova immediatamente che:

- $f(x)$  è normalizzata se  $c = \frac{1}{b-a}$
- il valore di aspettazione è  $E[X] = \int_a^b x \cdot c \, dx = \frac{a+b}{2}$
- la varianza è  $var(x) = \int_a^b x^2 \cdot c \, dx - E[X]^2 = \frac{(b-a)^2}{12}$

Questa funzione di distribuzione viene utilizzata in particolare quando mancano informazioni. Ad esempio, non caso in cui gli errori di misura dominanti siano dovuti alla risoluzione di lettura dello strumento (errore massimo  $\Delta X$ ) non si può sapere che valore la grandezza fisica ha all'interno dell'intervallo di ampiezza  $2\Delta X$  per cui si assume che il "valore vero" abbia una distribuzione uniforme nell'intervallo di ampiezza  $2\Delta X$  (in realtà è l'intervallo che ha una distribuzione uniforme...) e quindi si associa alla misura un'incertezza statistica  $\sigma_x = \frac{\Delta X}{\sqrt{3}}$ . In questo caso è  $P(\mu_x - \sigma_x \leq X \leq \mu_x + \sigma_x) = 57\%$ .

È interessante calcolare la funzione generatrice dei momenti e confrontare indici e coefficienti di forma con quelli della funzione di distribuzione normale.

Dalla loro definizione, si ricava facilmente che le funzioni generatrici dei momenti sono

$$M_x^*(t) = \frac{1}{t(b-a)} (e^{bt} - e^{at})$$
$$M_x(t) = \frac{1}{t(b-a)} (e^{(b-a)t/2} - e^{-(b-a)t/2}).$$

Calcolando le derivate delle funzioni generatrici rispetto a  $t$  in  $t = 0$ , e facendo attenzione alle forme indeterminate, si ottiene:

$\mu_0^* = 1$ ,  $\mu_1^* = (a + b)/2$ , come previsto,  
 $\mu_0 = 1$ ,  $m\mu_1 = 0$ ,  $\mu_2 = \sigma_x^2 = (b - a)^2/12$ , come previsto, e  
 $\mu_3 = 0$ ,  $\mu_4 = (b - a)^4/80$ , da cui si ricava il coefficiente di asimmetria  $\gamma_1 = 0$ , e in effetti la funzione di distribuzione è simmetrica rispetto al valore di aspettazione di  $x$  (come la funzione di distribuzione di Gauss). Per quanto riguarda la curtosi, è

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = -\frac{6}{5},$$

valore negativo per una funzione di distribuzione più "piatta" rispetto alla funzione di distribuzione di Gauss.

### 1.2.3 Funzione di distribuzione esponenziale

Anche questa funzione di distribuzione è molto usata in fisica: come si vedrà a proposito delle funzioni di Erlang, è la funzione di distribuzione dei tempi di attesa di eventi casuali indipendenti (decadimenti di nuclei radioattivi o di particelle instabili, ...). La variabile casuale continua può assumere solo valori positivi e la funzione di distribuzione è data da

$$f(x) = \frac{1}{\tau} e^{-x/\tau}$$

Risolvendo gli integrali, si vede immediatamente che:

- $f(x)$  è normalizzata
- il valore di aspettazione è  $\mu_x = \int_0^\infty x f(x) dx = \tau$
- la varianza è  $\sigma_x^2 = \int_0^\infty (x - \tau)^2 f(x) dx = \tau^2$

Anche in questo caso è interessante calcolare la funzione generatrice dei momenti e confrontare gli indici e coefficienti di forma con quelli della funzione di distribuzione normale. Usando la loro definizione, si ricava che le funzioni generatrici dei momenti sono

$$M_x^*(t) = \frac{1}{1 - t\tau}, \quad M_x(t) = \frac{e^{-t\tau}}{1 - t\tau}.$$

Calcolando le derivate delle funzioni generatrici rispetto a  $t$  in  $t = 0$ , con calcoli piuttosto

lunghi ma consigliati, si ottiene:

$\mu_0^* = 1$ ,  $\mu_1^* = \tau$ , come previsto,

$\mu_0 = 1$ ,  $\mu_1 = 0$ ,  $\mu_2 = \sigma_x^2 = \tau^2$ , come previsto, e

$\mu_3 = 2\tau^3$ ,  $\mu_4 = 9\tau^4$ , da cui si ricava il coefficiente di asimmetria  $\gamma_1 = 2$ , e in effetti la funzione di distribuzione è asimmetrica rispetto al valore di aspettazione di  $x$ . Per quanto riguarda il coefficiente di curtosi, è

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = 6,$$

positivo: funzione di distribuzione meno "piatta" rispetto alla funzione di distribuzione di Gauss.

Come già detto, su questa importante funzione di distribuzione si tornerà nella parte dedicata alle funzioni di Erlang.

### 1.2.4 Distribuzione binomiale

Se un fenomeno statistico può verificarsi soltanto in due modalità (favorevole o sfavorevole), e se la probabilità che il singolo evento sia favorevole è  $p$  (e la probabilità che sia sfavorevole è  $q = 1 - p$ ), la probabilità di avere  $k$  eventi favorevoli su  $n$  eventi indipendenti è data dalla distribuzione di probabilità binomiale:

$$P(k; n, p) = \binom{n}{k} p^k \cdot q^{n-k} \quad 0 \leq k \leq n$$

che dipende da due parametri,  $n$  e  $p$ .

Si verifica facilmente che:

- la condizione di normalizzazione è verificata:  $\sum_{k=0}^n \binom{n}{k} p^k \cdot q^{n-k} = 1$
- il valore di aspettazione è  $E[k] = \sum_{k=0}^n \binom{n}{k} k \cdot p^k \cdot q^{n-k} = np$
- la varianza è  $var(k) = E[k^2] - E[k]^2 = npq$

Le funzioni generatrici dei momenti sono:

$$\begin{aligned} M_k^*(t) &= E[e^{tk}] = \sum_{k=0}^n e^{tk} \binom{n}{k} \cdot p^k \cdot q^{n-k} = \sum_{k=0}^n \binom{n}{k} \cdot (e^t p)^k \cdot q^{n-k} = (pe^t + q)^n \\ M_x(t) &= e^{-t np} M_x^*(t) = (pe^{qt} + qe^{-pt})^n \end{aligned}$$

Calcolando le derivate di  $M_k^*(t)$  in  $t = 0$  si ottengono i momenti algebrici:

$$\mu_0^* = 1, \quad \mu_1^* = E[k] = np$$

e calcolando le derivate di  $M_k(t)$  in  $t = 0$  si ottengono i momenti centrali:

$$\begin{aligned} \mu_0 &= 1, \quad \mu_1 = 0, \quad \mu_2 = var(k) = npq, \\ \mu_3 &= npq(1 - 2p), \quad \mu_4 = npq[1 - 3pq(2 - n)]. \end{aligned}$$

Il coefficiente di asimmetria è quindi

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{1 - 2p}{\sqrt{npq}};$$

la distribuzione binomiale è quindi simmetrica (come la funzione di distribuzione di Gauss) per  $p = 0.5$ , e tende a diventarlo per  $n \rightarrow \infty$ .

Il coefficiente di curtosi è

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{1 - 6pq}{npq}$$

da cui si vede che la distribuzione binomiale tende a zero per  $n \rightarrow \infty$ .  
 Si conclude quindi che la distribuzione binomiale tende ad avere una forma simile alla funzione di distribuzione di Gauss per  $n \rightarrow \infty$ .

La distribuzione binomiale viene usata per descrivere il numero di eventi in un intervallo di un istogramma (come si vedrà dopo aver introdotto la distribuzione di Poisson), ma viene usata ogni volta in cui si hanno eventi indipendenti e solo due possibilità. Un tipico esempio è costituito dal numero di coppie di vaolri  $(x_i, y_i)$  di due variabili casuali che cadono all'interno di una regione nel piano  $xy$ .

### 1.2.5 Distribuzione di Poisson

La distribuzione di Poisson, o degli “eventi rari”, è matematicamente il limite della distribuzione binomiale quando il numero di eventi indipendenti  $n \rightarrow \infty$  e la probabilità che il singolo evento sia favorevole  $p \rightarrow 0$  in modo che  $E[k] = np = \nu$  resti costante. La probabilità di avere  $k$  eventi favorevoli se in media ce ne sono  $\nu$  è:

$$P_k = P(k; \nu) = \frac{\nu^k e^{-\nu}}{k!} \quad 0 \leq k \leq +\infty$$

La distribuzione di Poisson dipende da un solo parametro,  $\nu$ , ed ha una forma molto diversa a seconda del suo valore, come si può vedere in fig. 2.

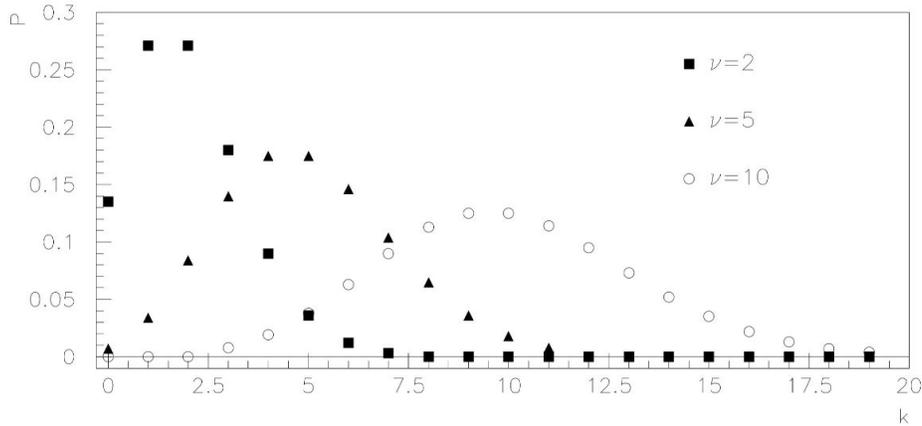


Figura 2: Distribuzione di Poisson per diversi valori di  $\nu$ .

È facile verificare che:

- la distribuzione è normalizzata:  $\sum_{k=0}^{+\infty} \nu^k \frac{e^{-\nu}}{k!} = e^{-\nu} \sum_{k=0}^{\infty} \frac{\nu^k}{k!} = 1$

- il valore di aspettazione è  $E[k] = \sum_{k=0}^{+\infty} k \cdot \nu^k \frac{e^{-\nu}}{k!} = e^{-\nu} \sum_{k=1}^{\infty} \frac{\nu^k}{(k-1)!} = \nu$

- la varianza è  $var(k) = E[k^2] - E[k]^2 = (\nu^2 + \nu) - \nu^2 = \nu$

e che le funzioni generatrici dei momenti sono:

$$M_k^*(t) = E[e^{tk}] = \sum_{k=0}^{+\infty} e^{tk} \cdot \nu^k \frac{e^{-\nu}}{k!} = e^{-\nu} \sum_{k=0}^{+\infty} \frac{(\nu e^t)^k}{k!} = e^{-\nu} e^{\nu e^t} = e^{\nu(e^t-1)}$$

$$M_k(t) = E[e^{t(k-\nu)}] = e^{-\nu t} M_k^*(t) = e^{-\nu t} e^{\nu(e^t-1)} = e^{\nu(e^t-t-1)}$$

come si ottiene si facilmente anche considerando il limite delle funzioni generatrici dei momenti della distribuzione binomiale per  $n \rightarrow \infty$  con  $p \rightarrow 0$  e  $np = \nu$  costante.

I momenti si ottengono al solito calcolando le derivate delle funzioni generatrici in  $t = 0$ :

$$\mu_0^* = M_k^*(t=0) = 1, \quad \mu_1^* = \left[ \frac{\partial M_k^*}{\partial t} \right]_{t=0} = E[k] = \nu$$

$$\mu_0 = M_k(t=0) = 1, \quad \mu_1 = \left[ \frac{\partial M_k}{\partial t} \right]_{t=0} = 0$$

$$\mu_2 = var(k) = \left[ \frac{\partial^2 M_k}{\partial t^2} \right]_{t=0} = \nu, \quad \mu_3 = \left[ \frac{\partial^3 M_k}{\partial t^3} \right]_{t=0} = \nu, \quad \mu_4 = \left[ \frac{\partial^4 M_k}{\partial t^4} \right]_{t=0} = 3\nu^2 + \nu$$

da cui

$$\gamma_1 = \frac{1}{\sqrt{\nu}}, \quad \gamma_2 = \frac{1}{\nu}$$

Come si vede, entrambi i coefficienti tendono a zero per  $\nu \rightarrow \infty$  e quindi anche la distribuzione di Poisson tende ad avere una forma simile a quella della funzione di distribuzione di Gauss quando il valore di aspettazione del numero di eventi è elevato.

L'applicazione tipica è relativa ai conteggi in un intervallo di tempo fissato. Supponiamo di avere in media  $\nu$  eventi indipendenti nell'unità di tempo e quindi in media  $\nu \cdot \Delta t$  eventi nell'intervallo di tempo  $\Delta t$ . La variabile casuale  $k$ , numero di eventi in  $\Delta t$ , ha distribuzione di Poisson. Infatti, possiamo dividere l'intervallo  $\Delta t$  in  $n$  intervalli di ampiezza  $\delta t = \Delta t/n$  sufficientemente piccola, in modo tale che la probabilità di avere due o più eventi nello stesso intervallo sia trascurabile e che la probabilità di avere un evento sia proporzionale a  $\delta t$  ( $p = \nu \cdot \delta t$ ). In ciascuno degli  $n$  intervalli si potranno quindi avere solo 0 o 1 eventi e la probabilità di avere  $k$  intervalli su  $n$  con 1 evento è data dalla distribuzione binomiale con probabilità di evento favorevole  $p$  e valore di aspettazione di  $k$   $E[k] = \nu \cdot \Delta t$ . Al limite per  $n \rightarrow \infty$  e  $p \rightarrow 0$  si ottiene la distribuzione di Poisson.

### 1.2.6 Istogrammi

La distribuzione binomiale e la distribuzione di Poisson vengono usate anche per determinare l'errore statistico del numero di eventi in un intervallo di un istogramma. Dato un campione di  $n$  misure indipendenti di una grandezza fisica  $x$ , se si considera il generico intervallo  $k$ , si può assumere che il numero  $n_k$  di eventi in quell'intervallo sia una variabile casuale che segue una distribuzione binomiale. Infatti possiamo dire che ci sono solo due possibilità: o un valore cade in quell'intervallo (caso favorevole, probabilità  $p_k$ ), oppure cade fuori (sfavorevole, probabilità  $q_k = 1 - p_k$ )<sup>4</sup>. Il valore di aspettazione di  $n_k$  è  $np_k$  e la deviazione standard è  $\sigma_{n_k} = \sqrt{np_kq_k}$ . Se la funzione di distribuzione  $f(x)$  è nota, la probabilità che il singolo evento sia favorevole si ottiene integrando  $f(x)$  sul  $k$ -esimo intervallo. Se  $f(x)$  non è nota, ai fini del calcolo dell'incertezza statistica, si assume  $p_k \simeq n_k^m/n$ , dove  $n_k^m$  è il numero di eventi misurato nell'intervallo  $k$ , e quindi  $\sigma_{n_k} = \sqrt{n_k^m(1 - n_k^m/n)}$ . Più semplice il caso in cui il numero di valori  $n$  è molto grande e le probabilità  $p_k$  relative ai diversi intervalli sono tutte molto piccole (minori del 10%): in questo caso si può assumere che i numeri di eventi nei diversi istogrammi  $n_j$  abbiano distribuzione di Poisson, con valore di aspettazione e varianza uguali a  $n_j^m$ , cioè si può usare come errore statistico semplicemente la radice quadrata del numero di eventi misurati.

---

<sup>4</sup>in realtà questo non è formalmente corretto: vedi distribuzione multinomiale.

### 1.2.7 Distribuzione Gamma

La distribuzione Gamma è una funzione di distribuzione che comprende come casi particolari le funzioni di distribuzione di Erlang e di  $\chi^2$ . Viene utilizzata come modello generale dei tempi di attesa nella “teoria delle code”.

Dipende da due parametri reali positivi  $\alpha$ ,  $\beta$  ed è data da:

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad x \geq 0$$

dove  $\Gamma(\alpha)$  è la funzione *Gamma di Eulero* che è definita come

$$\Gamma(\alpha) = \int_0^{+\infty} y^{\alpha-1} e^{-y} dy$$

e che, come si verifica facilmente, soddisfa le seguenti relazioni:

$$\begin{aligned} \Gamma(1) &= \int_0^{+\infty} e^{-y} dy = 1, & \Gamma(1/2) &= \int_0^{+\infty} y^{-\frac{1}{2}} e^{-y} dy = \sqrt{\pi} \\ \Gamma(\alpha + 1) &= \int_0^{+\infty} y^\alpha e^{-y} dy = \alpha \Gamma(\alpha), & \Gamma(n) &= (n-1)! \quad n \in \mathbb{N} \end{aligned}$$

Usando la sostituzione  $y = x/\beta$  e la definizione della funzione *Gamma* si può dimostrare che:

- la distribuzione Gamma è normalizzata:  $\int_0^{+\infty} f(x; \alpha, \beta) dx = 1$
- il valore di aspettazione è  $E[x] = \int_0^{+\infty} x \cdot f(x; \alpha, \beta) dx = \alpha\beta$
- la varianza è  $var(x) = \alpha\beta^2$

Infine, le funzioni generatrici dei momenti sono:

$$M_x^*(t) = (1 - \beta t)^{-\alpha}, \quad M_x(t) = e^{-t\alpha\beta} (1 - \beta t)^{-\alpha}.$$

La distribuzione Gamma, con i suoi due parametri, è piuttosto complicata. Qui ci limitiamo ai due casi particolari costituiti dalla funzione di distribuzione di Erlang e dalla la funzione di distribuzione di  $\chi^2$ .

### 1.2.8 Distribuzione di Erlang

Si ottiene dalla distribuzione Gamma ponendo  $\alpha = k$  con  $k$  intero e  $\beta = 1$ . La sua espressione è:

$$g(x, k) = \frac{1}{(k-1)!} x^{k-1} e^{-x}, \quad x \geq 0$$

ed è

- valore di aspettazione:  $E[X] = \frac{1}{(k-1)!} \int_0^{+\infty} x^k e^{-x} dx = \frac{\Gamma(k+1)}{(k-1)!} = k$

- varianza:  $var(X) = E[X^2] - E[X]^2 = \frac{1}{(k-1)!} \int_0^{+\infty} x^{k+1} e^{-x} dx - k^2 = (k+1)k - k^2 = k$

in accordo con valore di aspettazione e varianza della distribuzione Gamma.

Le funzioni di Erlang sono le funzioni di distribuzione dei tempi di attesa di eventi casuali. Mentre la distribuzione di Poisson dà la probabilità che accadano  $n$  eventi indipendenti in un certo intervallo di tempo fisso quando in media ce ne sono  $\nu$ , la distribuzione di Erlang dà la distribuzione del tempo che intercorre tra un istante arbitrario e il primo, il secondo ... il  $k$ -esimo evento (o tra un evento casuale e il primo, il secondo ... il  $k$ -esimo evento successivo). La funzione di distribuzione del tempo di attesa  $t$  del  $k$ -esimo evento è infatti

$$f(t; k) = \mu \frac{(\mu t)^{k-1}}{(k-1)!} e^{-\mu t},$$

dove  $\mu$  è il numero medio di eventi nell'unità di tempo, che si ottiene dalla funzione di Erlang per la variabile  $t = x/\mu$ .

\* Dimostrazione

Ricaviamo la distribuzione del tempo di attesa del generico evento  $k$ . Essendo gli eventi indipendenti, la probabilità  $P_k(t, t + \delta t)$  che il  $k$ -esimo evento cada nell'intervallo di tempo  $(t, t + \delta t)$  è data dal prodotto della probabilità  $P_{k-1}(0, t)$  di avere  $k-1$  eventi in  $(0, t)$  e della probabilità  $P_1(t, t + \delta t)$  di avere un evento in  $(t, t + \delta t)$ :

$$P_k(t, t + \delta t) = P_{k-1}(0, t) \cdot P_1(t, t + \delta t).$$

Per calcolare  $P_1(t, t + \delta t)$ , bisogna ricordare che si parla di eventi indipendenti e che quindi avere o meno un evento in  $(t, t + \delta t)$  deve essere indipendente da quanto successo prima; in altre parole,  $P_1(t, t + \delta t)$  non può dipendere da  $t$  ma deve dipendere solo dall'ampiezza  $\delta t$  dell'intervallo. Assumiamo poi che questa ampiezza sia sufficientemente piccola da poter trascurare la probabilità di avere due o più eventi in  $\delta t$  sia trascurabile e che la probabilità di averne uno sia proporzionale a  $\delta t$  stesso. Si ottiene quindi  $P_1(t, t + \delta t) = \mu \delta t$ , con  $\mu$  numero medio di eventi nell'unità di tempo.

La probabilità di avere  $k-1$  eventi  $(0, t)$  (quando in media ce ne sono  $\nu = \mu t$ ) è data dalla distribuzione di Poisson:

$$P_{k-1}(0, t) = \nu^{k-1} \frac{e^{-\nu}}{(k-1)!} = (\mu t)^{k-1} \frac{e^{-\mu t}}{(k-1)!}.$$

Si ottiene quindi:

$$P_k(t, t + \delta t) = \frac{(\mu t)^{k-1}}{(k-1)!} e^{-\mu t} \mu \delta t$$

e, per definizione di funzione di distribuzione, la funzione di distribuzione dei tempi di attesa del  $k$ -esimo evento è:

$$f(t; k) = \frac{P_k(t, t + \delta t)}{\delta t} = \mu \frac{(\mu t)^{k-1}}{(k-1)!} e^{-\mu t}$$

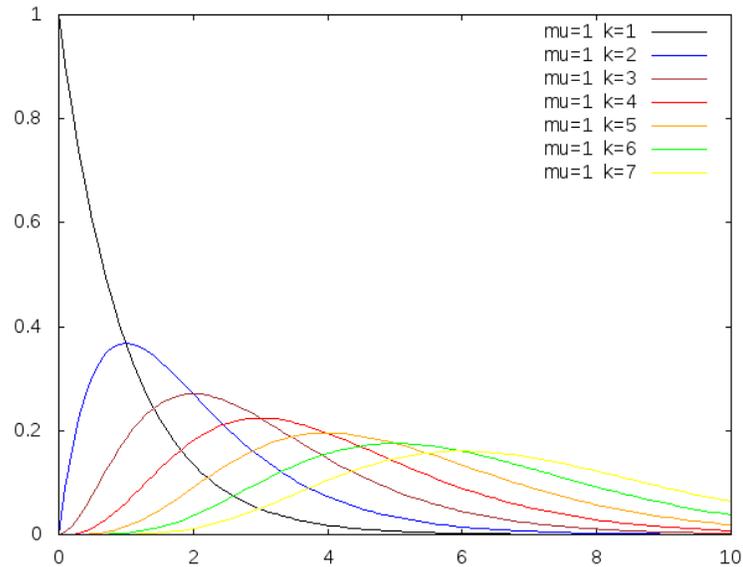


Figura 3: Funzione di distribuzione di Erlang per diversi valori di  $k, \mu$ .

Dalla funzione di distribuzione dei tempi di attesa  $f(t, k)$  (ma anche da quanto visto per le funzione di Erlang) si ottiene facilmente che è:

$$E[t] = \frac{k}{\mu}, \quad \sigma_t^2 = \frac{k}{\mu^2}.$$

Non solo il valore di aspettazione e la varianza scalano con  $k$ , ma la forma della funzione di distribuzione cambia con  $k$ , in particolare per valori piccoli del parametro, come si vede dalla fig. 3 e dalle espressioni esplicite delle funzioni:

$$\begin{aligned} f(t; k = 1) &= \mu e^{-\mu t} \\ f(t; k = 2) &= \mu^2 t e^{-\mu t} \\ f(t; k = 3) &= \frac{\mu^3 t^2}{2} e^{-\mu t} \\ &\dots \end{aligned}$$

Il tempo di attesa per il primo evento ( $k = 1$ ) ha una funzione di distribuzione esponenziale negativa, tipica dei decadimenti di nuclei e particelle instabili. In questi casi, si introduce la "vita media" definita come il valore di aspettazione di  $t$ :  $\tau = 1/\mu$ . La derivazione della distribuzione dei tempi di attesa in questo caso, fatta senza usare la distribuzione di Poisson, si può trovare in appendice.

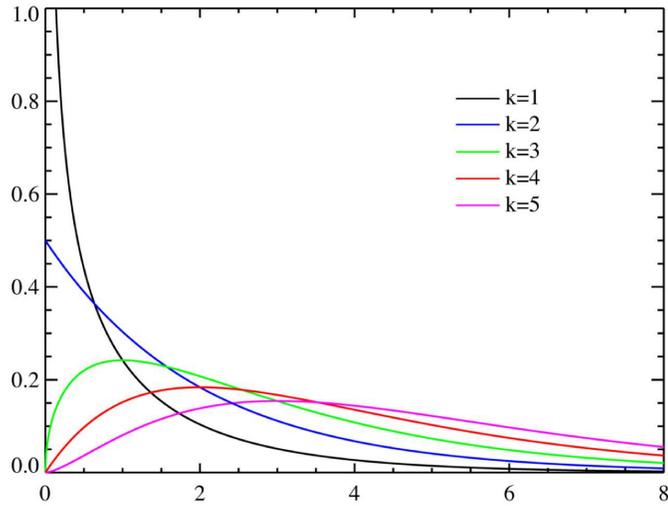


Figura 4: Funzione di distribuzione di  $\chi^2$  per diversi valori di  $n$ .

### 1.2.9 Distribuzione di $\chi^2$

La funzione di distribuzione di  $\chi^2$  si ottiene dalla distribuzione Gamma per  $\alpha = n/2$ , con  $n$  intero maggiore di zero, e  $\beta = 2$ . Dipende quindi dal solo parametro  $n$  (spesso indicato con  $\nu$  e detto *numero di gradi di libertà*). La sua espressione è:

$$f(x; n) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad x \geq 0$$

È molto diffusa in quanto è la funzione di distribuzione di variabili casuali definite come somme dei quadrati di  $n$  variabili casuali indipendenti con funzione di distribuzione normale standard (vedi cap. ...), e la incontreremo nel seguito, anche per i test d'ipotesi.

Valore di aspettazione e varianza si possono ottenere da quelli della distribuzione Gamma o calcolare direttamente:

$$\begin{aligned} \mu_x &= E[x] = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_0^{+\infty} x^{\frac{n}{2}} e^{-\frac{x}{2}} dx = n \\ \sigma_x^2 &= var(x) = E[x^2] - E[x]^2 = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_0^{+\infty} x^{\frac{n}{2}+1} e^{-\frac{x}{2}} dx - n^2 = 2n \end{aligned}$$

È interessante notare come l'andamento della funzione di distribuzione di  $\chi^2$  cambi con  $n$  per  $n$  piccoli. Di seguito si riporta la forma esplicita della distribuzione di  $\chi^2$  per i primi

5 gradi di libertà:

$$\begin{aligned}\chi_1^2 &: \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\frac{x}{2}} \\ \chi_2^2 &: \frac{1}{2} e^{-\frac{x}{2}} \\ \chi_3^2 &: \frac{1}{\sqrt{2\pi}} \sqrt{x} e^{-\frac{x}{2}} \\ \chi_4^2 &: \frac{1}{4} x e^{-\frac{x}{2}} \\ \chi_5^2 &: \frac{1}{3\sqrt{2\pi}} x\sqrt{x} e^{-\frac{x}{2}}\end{aligned}$$

Come si vede anche da fig. 4, per  $n = 1$  la funzione diverge per  $x = 0$ , mentre per  $n = 2$  è esponenziale, e per  $n \geq 3$  è uguale a zero per  $x = 0$ .

Le funzioni generatrici dei momenti sono:

$$\begin{aligned}M_x^*(t) &= E[e^{tx}] = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \int_0^{+\infty} e^{-\frac{x}{2}(1-2t)} x^{\frac{n}{2}-1} dx = (1-2t)^{-\frac{n}{2}} \\ M_x(t) &= e^{-\mu_x t} M_x^*(t) = e^{-tn} (1-2t)^{-\frac{n}{2}}\end{aligned}$$

\* Per ricavare  $M_x^*(t)$  basta effettuare la sostituzione  $y = x(1-2t)$ , e usare la normalizzazione della funzione di distribuzione Gamma.

Calcolando le loro derivate in  $t = 0$ , si ottengono i momenti algebrici e centrali:

$$\begin{aligned}\mu_0^* &= M_x^*(t=0) = 1, \quad \mu_1^* = \left[ \frac{\partial M_x^*}{\partial t} \right]_{t=0} = E[x] = n \\ \mu_0 &= M_x(t=0) = 1, \quad \mu_1 = \left[ \frac{\partial M_x}{\partial t} \right]_{t=0} = 0, \quad \mu_2 = var(x) = \left[ \frac{\partial^2 M_x}{\partial t^2} \right]_{t=0} = 2n \\ \mu_3 &= \left[ \frac{\partial^3 M_x}{\partial t^3} \right]_{t=0} = 8n, \quad \mu_4 = 12n^2 + 48n.\end{aligned}$$

Da questi ultimi si ottengono i coefficienti di asimmetria e di curtosi

$$\gamma_1 = \frac{2\sqrt{2}}{\sqrt{n}}, \quad \gamma_2 = \frac{12}{n}.$$

che tendono a 0 per  $n \rightarrow \infty$ : questo significa che la funzione tende a diventare simmetrica e che la piatezza della distribuzione tende a quella della distribuzione normale.

In modo più rigoroso, si può dimostrare che le funzioni generatrici dei momenti tendono, per  $n \rightarrow \infty$ , tendono alle funzioni generatrici della distribuzione di Gauss.

\* Dimostrare che per  $n \rightarrow \infty$   $x$  ha funzione di distribuzione di Gauss equivale a dimostrare che  $y = (x-n)/\sqrt{2n}$  tende ad avere distribuzione normale standard. Per dimostrare questo si scrive la funzione generatrice dei momenti di  $y$ :

$$M_y^*(t) = E[e^{t(x-n)/\sqrt{2n}}] = e^{-t\sqrt{n/2}} E[e^{tx/\sqrt{2n}}] = e^{-t\sqrt{n/2}} M_x^*(t/\sqrt{2n}) = e^{-t\sqrt{n/2}} (1-t\sqrt{2/n})^{-n/2}.$$

Si calcola quindi il suo logaritmo, lo si sviluppa in serie in un intorno di  $t\sqrt{2/n} = 0$  e si calcola il suo limite per  $n \rightarrow \infty$  ottenendo  $t^2/2$ . Quindi, per  $n \rightarrow \infty$ , la funzione generatrice dei momenti coincide con la funzione generatrice dei momenti della funzione di distribuzione normale standard.

In pratica questo è vero per  $n$  relativamente grandi, dell'ordine di 100, per cui le funzioni di ripartizione si trovano tabulate in genere fino a un numero di gradi di libertà pari a 100, e si rimanda a quelle della distribuzione normale standard per  $n$  maggiori. Cosa piuttosto utile in pratica per il test d'ipotesi.

## Distribuzione di Maxwell-Boltzmann

Un'applicazione interessante è l'uso della funzione di distribuzione di  $\chi^2$  per ricavare la distribuzione di Maxwell-Boltzmann, la funzione di distribuzione delle velocità molecolari di un gas perfetto in meccanica classica (termodinamica statistica, teoria cinetica dei gas). Consideriamo un gas perfetto (o ideale) costituito da molecole puntiformi che interagiscono solo tramite urti elastici, che non sono soggette a forze di interazione a distanza, che la loro massa è tale da poter trascurare la forza di gravità. La posizione delle molecole nello spazio è uniforme e la loro direzione casuale (non esiste una direzione privilegiata).

Introdotta un sistema di riferimento ortogonale  $x, y, z$  (1, 2, 3), indichiamo con  $\vec{v} = (v_1, v_2, v_3)$  la velocità di una generica molecola. Nelle ipotesi fatte, le componenti della velocità sono variabili casuali indipendenti, che, non essendoci direzioni privilegiate, devono avere valore di aspettazione zero e tutte la stessa varianza  $\sigma^2$ . L'ipotesi più semplice è che abbiano distribuzione normale  $N(0, \sigma^2)$ .

Tuttavia di interesse non è la distribuzione delle componenti ma quella dei moduli delle velocità:  $v = \sqrt{\sum_i v_i^2}$  con  $i = 1, 2, 3$ .

Per trovarla, introduciamo  $\vec{q} = \frac{1}{\sigma} \vec{v}$ , le cui componenti  $q_i$  hanno tutte distribuzione  $N(0,1)$ . Quindi la variabile casuale  $z = \sum_i q_i^2$ ,  $i = 1, 2, 3$ , è la somma dei quadrati di 3 variabili casuali indipendenti con distribuzione normale standard, e ha funzione di distribuzione di  $\chi^2$  a 3 gradi di libertà:

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

Dalla definizione di  $z$  otteniamo  $v = \sigma\sqrt{z}$ , e quindi:

$$g(v) = f(z(v)) \left| \frac{dz}{dv} \right| = \sqrt{\frac{2}{\pi}} \frac{v^2}{\sigma^3} e^{-\frac{v^2}{2\sigma^2}}$$

Ponendo  $\sigma^2 = \frac{k_B T}{m}$  otteniamo la forma "tradizionale" della funzione di distribuzione delle velocità molecolari. Da questa si ottiene la velocità media

$$E[v] = \bar{v} = \int_0^\infty g(v) dv = \sqrt{\frac{8kT}{\pi m}}$$

ma anche la velocità più probabile, le velocità quadratiche media...

### 1.3 Più variabili casuali

Spesso si ha a che fare con problemi in cui intervengono più variabili casuali che devono essere considerate contemporaneamente. Un esempio tipico è in caso di un esperimento in cui vengono eseguite  $n$  misure di una grandezza fisica  $X$ : i risultati delle misure possono essere visti  $n$  valori di una variabile casuale  $x$  con distribuzione  $f(x)$ , ma anche come i valori di  $n$  variabili casuali  $x_i$  tutte con funzione di distribuzione  $f(x)$ . L'insieme delle  $n$  variabili casuali viene detto campione casuale (o aleatorio) di dimensione  $n$ . Altro esempio sono le stime di parametri diversi ottenute dallo stesso set di dati.

In questa parte, oltre alla trattazione generale, vengono brevemente descritte due distribuzioni: la distribuzione multinomiale e la funzione di distribuzione multinormale (o normale multivariata).

#### 1.3.1 Funzione di distribuzione congiunta, marginale e condizionata

Nel caso generale di  $n$  variabili casuali  $x_i$ , ognuna definita nel rispettivo spazio  $\Omega_i$ , si introducono:

- la funzione di distribuzione congiunta  $f(x_1, \dots, x_n)$ :

è la funzione delle  $n$  variabili casuali  $x_i$  che permette di calcolare la probabilità che queste cadano contemporaneamente in determinati intervalli  $A_i \subset \Omega_i$ :

$$P(x_1 \in A_1, \dots, x_n \in A_n) = \int_{A_1} \dots \int_{A_n} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Considerando l'intero spazio campionario  $\vec{\Omega}_x$ , deve essere:

$$P(x_1 \in \Omega_1, \dots, x_n \in \Omega_n) = \int_{\Omega_1} \dots \int_{\Omega_n} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$$

(condizione di normalizzazione). Come la funzione di distribuzione di una variabile casuale, sempre per definizione di probabilità, la funzione di distribuzione congiunta deve anche essere definita positiva e limitata in  $\vec{\Omega}_x$ .

Nel caso particolare di variabili casuali indipendenti, essa è ovviamente definita come il prodotto delle funzioni di distribuzione  $f_i$  delle singole variabili:

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \dots \cdot f_n(x_n)$$

- le funzioni di distribuzione marginali:

solo le funzioni di distribuzione delle diverse variabili casuali  $x_i$  valutate indipendentemente dalle (o meglio, integrando sui possibili valori delle) altre  $n - 1$  variabili:

$$f_{M_i}(x_i) = \int_{\Omega_1} \dots \int_{\Omega_{i-1}} \int_{\Omega_{i+1}} \dots \int_{\Omega_n} f(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

dove l'integrazione avviene su tutte le  $x_j \neq x_i$ . Coi, la funzione marginale di  $x_1$  è:

$$f_{M_1}(x_1) = \int_{\Omega_2} \dots \int_{\Omega_n} f(x_1, \dots, x_n) dx_2 \dots dx_n$$

Come dovrebbe essere chiaro dalla definizione, le funzioni di distribuzione marginali sono automaticamente normalizzate.

- le funzioni di distribuzione condizionate:

sono le funzioni di distribuzione di una variabile  $x_i$  mentre le altre assumono valori fissati.

Considerando  $i = 1$ , per semplicità di scrittura, la funzione di distribuzione condizionata della variabile  $x_1$  per  $x_j = x_j^*$ ,  $j = 2, \dots, n$  è:

$$f_{C_1}(x_1) = g(x_1 | x_2^*, \dots, x_n^*) = c \cdot f(x_1, x_2^*, \dots, x_n^*)$$

dove  $c$  è una costante necessaria affinché  $f_{C_1}(x_1)$  sia normalizzata, ed è:

$$c = \left( \int_{\Omega_1} f(x_1, x_2^*, \dots, x_n^*) dx_1 \right)^{-1}$$

È importante sottolineare che, nel caso di variabili casuali indipendenti, si ha  $f_i(x_i) = f_{M_i}(x_i) = f_C(x_i)$

### 1.3.2 Valore di aspettazione e varianza

Valori di aspettazione e varianze delle variabili casuali  $x_i$  con densità di probabilità congiunta  $f(x_1, \dots, x_n)$ , sono dati da:

$$E[x_i] = \mu_i = \int_{\Omega_1} \dots \int_{\Omega_n} x_i f(x_1, \dots, x_n) dx_1 \dots dx_n$$

$$var(x_i) = \sigma_i^2 = E[(x_i - \mu_i)^2] = E[x_i^2] - E[x_i]^2 = \int_{\Omega_1} \dots \int_{\Omega_n} (x_i - \mu_i)^2 f(x_1, \dots, x_n) dx_1 \dots dx_n$$

ed è ancora  $var(x_i) = E[x_i^2] - E[x_i]^2$ .

### 1.3.3 Covarianza e coefficiente di correlazione

Un parametro importante nel caso di più variabili casuali è la covarianza. La covarianza fra due variabile  $x_i$  e  $x_j$  è definita come:

$$\begin{aligned} cov(x_i, x_j) &= E[(x_i - \mu_i)(x_j - \mu_j)] = \int_{\Omega_1} \dots \int_{\Omega_n} (x_i - \mu_i)(x_j - \mu_j) f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= E[x_{ij}] - \mu_i \mu_j \end{aligned}$$

Se  $i = j$  è  $cov(x_i, x_i) = \sigma_i^2$ .

La covarianza indica la correlazione fra le due variabili casuali. Può essere positiva o negativa e in genere è diversa da 0. Come si vede dalla sua definizione, è uguale a zero per variabili indipendenti (diverse), cioè quando la densità di probabilità congiunta è il prodotto delle densità di probabilità delle variabili  $x_i$ , infatti in questo caso è  $E[x_{ij}] = E[x_i]E[x_j]$ .

Quindi per  $i \neq j$

$$cov(x_i, x_j) \begin{cases} = 0 & \text{se le } x_i \text{ sono indipendenti,} \\ \neq 0 & \text{se le } x_i \text{ sono correlate.} \end{cases}$$

Per esprimere il livello di correlazione tra due variabili casuali, è molto utile introdurre il coefficiente di correlazione:

$$\rho_{ij} = \frac{cov(x_i, x_j)}{\sigma_i \sigma_j}$$

che è sempre limitato tra  $-1$  e  $1$ , cioè

$$-1 \leq \rho_{xy} \leq +1$$

\* Dimostrazione:

Date due variabili casuali  $x$  e  $y$ , si definiscono le variabili casuali  $u = x - \mu_x$  e  $v = y - \mu_y$  e si introduce  $z = au - v$ . Si calcola il valore di aspettazione di  $z^2$ :

$$\begin{aligned} E[z^2] &= E[(au - v)^2] = a^2 \cdot E[u^2] + E[v^2] - 2a \cdot E[u \cdot v] = \\ &= a^2 \cdot E[(x - \mu_x)^2] + E[(y - \mu_y)^2] - 2a \cdot E[(x - \mu_x) \cdot (y - \mu_y)] = a^2 \cdot \sigma_x^2 + \sigma_y^2 - 2a \cdot cov(x, y) \end{aligned}$$

$E[z^2] \geq 0$  solo se il suo discriminante  $\Delta \leq 0$ , ovvero se  $\Delta = 4 \cdot (cov(x, y))^2 - 4 \cdot \sigma_x^2 \sigma_y^2 \leq 0$ :

$$(cov(x, y))^2 = (\rho_{xy})^2 \cdot \sigma_x^2 \sigma_y^2 \leq \sigma_x^2 \sigma_y^2 \quad \Rightarrow \quad |\rho_{xy}| \leq 1$$

Se  $\rho_{ij} = \pm 1$  allora le variabili casuali  $x_i$  e  $x_j$  sono completamente correlate, positivamente o negativamente: questo è il caso di variabili legate da una relazione lineare.

Se  $\rho_{ij} = 0$  allora le variabili sono scorrelate. Questa è una condizione necessaria ma non sufficiente perchè le variabili siano indipendenti.

Infine, in genere si dice che la correlazione è forte se  $0.7 < |\rho_{ij}| \leq 1$ , debole se  $|\rho_{ij}| < 0.3$ .

Varianze e covarianze sono gli elementi della matrice delle covarianze  $V$ :

$$\begin{aligned} V_{ij} &= \text{cov}(x_i, x_j) = \rho_{ij}\sigma_i\sigma_j \quad i \neq j \\ V_{ii} &= \text{var}(x_i) \end{aligned}$$

Quindi

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{21}\sigma_2\sigma_1 & \cdots & \rho_{n1}\sigma_n\sigma_1 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{n2}\sigma_n\sigma_2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}$$

è una matrice simmetrica e in particolare è diagonale se le variabili sono indipendenti.

Come vedremo, la matrice delle covarianze permette di semplificare notevolmente la notazione nel caso di più variabili casuali.

### 1.3.4 Esempi di variabili casuali correlate

Benchè il caso di variabili casuali indipendenti sia particolarmente semplice, spesso si ha a che fare con variabili correlate. Un esempio è costituito dai valori dei parametri stimati a partire da uno stesso insieme di dati. Lo studio delle distribuzioni delle stime fa parte della parte pratica del corso.

Un altro esempio, trattato in appendice B (cap. 7), è costituito da misure di grandezze fisiche affette da incertezze sistematiche. Si ricorda che per incertezza sistematica (da non confondere con l'errore sistematico) si intende l'incertezza che si ha dopo aver corretto al meglio delle conoscenze per gli effetti noti. In altre parole, è l'incertezza sulla correzione applicata o l'incertezza che si ha sul fatto che la correzione da applicare sia zero. E questa incertezza introduce correlazioni tra le misure, come abbastanza intuitivo.

### 1.3.5 Distribuzione multinomiale

Di seguito vengono trattate solamente due distribuzioni particolarmente utili per gli scopi del corso: la distribuzione di probabilità multinomiale (variabili casuali discrete) e, nel prossimo paragrafo, la distribuzione multinormale (variabili casuali continue).

Nel caso della distribuzione binomiale si considerano  $n$  eventi indipendenti, che possono essere solo “favorevoli” o “sfavorevoli”. Se  $p$  è la probabilità che il singolo evento sia favorevole, la probabilità di avere  $k$  eventi favorevoli su  $n$  è data da

$$P(k; n, p) = \binom{n}{k} p^k \cdot q^{n-k} \quad 0 \leq k \leq n$$

(vedi 1.2.4).

Se le alternative sono  $m > 2$ , con probabilità  $p_i$  ( $i = 1, m$  che il singolo evento si verifichi secondo la modalità  $i$  (e dovrà essere  $\sum_{j=1}^m p_j = 1$ ), la probabilità di avere, su  $n$  eventi,  $k_1$  eventi di modalità 1,  $k_2$  eventi di modalità 2, ...  $k_m$  eventi di modalità  $m$  (con  $\sum_{j=1}^m k_j = n$ ) è

$$P(k_1, \dots, k_m) = \frac{n!}{k_1! \dots k_m!} p_1^{k_1} \dots p_m^{k_m}.$$

La dimostrazione è simile a quella del caso della distribuzione binomiale, tenendo conto che ora il numero delle possibili sequenze è  $n!/(k_1! \dots k_m!)$ .

Valori di aspettazione e varianze sono

$$E[k_i] = np_i \quad \sigma_{k_i}^2 = np_i(1 - p_i)$$

come si può vedere abbastanza facilmente, e come per la distribuzione binomiale. Del resto, basta pensare che si possono sempre raggruppare tutte le modalità  $j \neq i$  in un'unica modalità, riconducendosi al caso della distribuzione binomiale.

In più, però, bisogna tener conto del fatto che le variabili  $k_i$  sono correlate, ed è

$$\text{cov}(k_i, k_j) = E[(k_i - np_i)(k_j - np_j)] = -np_i p_j,$$

diversa da zero e negativa. È quindi

$$\rho_{ij} = - \left[ \frac{p_i p_j}{(1 - p_i)(1 - p_j)} \right]^{\frac{1}{2}}.$$

\* Calcolo della covarianza, nel caso di tre sole modalità:

$$P(k_1, k_2, k_3) = \frac{n!}{k_1! k_2! k_3!} p_1^{k_1} p_2^{k_2} (1 - p_1 - p_2)^{n - k_1 - k_2}$$

$$\begin{aligned}
E[k_1 k_2] &= n(n-1)p_1 p_2 \sum_{k_1, k_2=1, n} \frac{(n-2)!}{(k_1-1)!(k_2-1)!(n-k_1-k_2)!} p_1^{k_1-1} p_2^{k_2-1} (1-p_1-p_2)^{n-k_1-k_2} \\
&= n(n-1)p_1 p_2 \\
cov(k_1, k_2) &= E[k_1 k_2] - E[k_1]E[k_2] = n^2 p_1 p_2 - n p_1 p_2 + n^2 p_1 p_2 = -n p_1 p_2
\end{aligned}$$

A rigore, il numero di eventi in un intervallo di un istogramma ha una distribuzione multinomiale e non binomiale: il singolo evento, se non cade in uno specifico intervallo, deve necessariamente cadere in uno degli altri  $m-1$  intervalli. I numeri di eventi nei diversi intervalli sono quindi correlati, e negativamente. Cosa di cui bisogna tener conto in particolare nella stima dei parametri e nel test d'ipotesi. Se però il numero di intervalli con eventi è elevato, come spesso accade, è  $p_i \ll 1$  per ogni  $i$  e si ottiene:  $E[k_i] = n p_i$ ,  $\sigma_{k_i}^2 \simeq n p_i$ , come nel caso della distribuzione di Poisson, e  $cov(k_i, k_j) \simeq 0$ . Se le probabilità elementari sono al massimo 0.1, infatti, il coefficiente di correlazione è minore di -0.11, e quindi la correlazione è debole, trascurabile in prima approssimazione.

Quanto ricavato è corretto se il numero totale di eventi  $n$  ha un valore ben definito, fissato "a priori", ma non nel caso in cui  $n$  sia una variabile casuale con distribuzione di Poisson e valore di aspettazione  $\nu$  (ad esempio, invece di contare  $n=100$  eventi, contiamo gli eventi in un intervallo di tempo in cui in media ce ne sono  $\nu=100$ ):

$$P(n) = \frac{e^{-\nu} \nu^n}{n!}.$$

La probabilità di avere  $n$  eventi di cui  $k_1$  secondo la modalità 1 ecc. è quindi il prodotto delle probabilità:

$$\begin{aligned}
P(n, k_1, \dots, k_m) &= \frac{e^{-\nu} \nu^n}{n!} \frac{n!}{k_1! \dots k_m!} p_1^{k_1} \dots p_m^{k_m} \\
&= \frac{e^{-\nu p_1}}{k_1!} (\nu p_1)^{k_1} \dots \frac{e^{-\nu p_m}}{k_m!} (\nu p_m)^{k_m}
\end{aligned}$$

come si ottiene usando  $\sum_j p_j = 1$  e  $\sum_j k_j = n$ . La distribuzione di probabilità congiunta è quindi il prodotto di  $m$  distribuzioni di Poisson con valori di aspettazione  $E[k_i] = \nu p_i$  e i numeri di eventi appartenenti alle diverse modalità sono variabili casuali indipendenti.

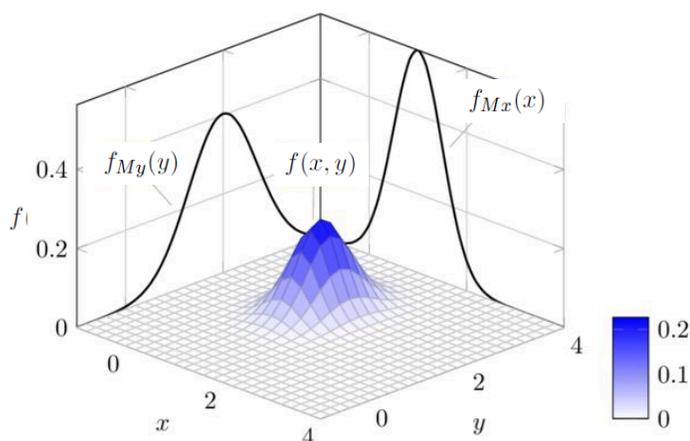


Figura 5: Esempio di distribuzione binormale con distribuzioni marginali (Raghavendra Selvan).

### 1.3.6 Distribuzione multinormale

La funzione di distribuzione multinormale (o normale multivariata) è l'estensione al caso a più dimensioni della distribuzione normale. È una funzione di distribuzione fondamentale, che verrà introdotta prima per due e più variabili casuali indipendenti, passando poi al caso più generale.

La funzione di distribuzione congiunta di due variabili casuali  $x_1$  e  $x_2$  con funzioni di distribuzione di Gauss  $f_1(x_1) = N(\mu_1, \sigma_1^2)$  e  $f_2(x_2) = N(\mu_2, \sigma_2^2)$  è, nel caso di variabili indipendenti, il prodotto delle due funzioni di distribuzione:

$$f(x_1, x_2) = f_1(x_1)f_2(x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-Q^2/2}$$

dove

$$Q^2 = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}.$$

Le funzioni di distribuzione marginali e condizionate sono

$$f_{M1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2, \quad f_{M2}(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_1$$

$$f_{C1}(x_1, x_2^*) = \frac{f(x_1, x_2^*)}{f_{M2}(x_2^*)}, \quad f_{C2}(x_2, x_1^*) = \frac{f(x_1^*, x_2)}{f_{M1}(x_1^*)}$$

e corrispondono rispettivamente alle proiezioni sul piano  $x_1, z$  o  $x_2, z$  e alle intersezioni (riscalate) della superficie  $f(x_1, x_2)$  con piani paralleli al piano  $x_2, z$  o  $x_1, z$ . Un esempio di funzioni di distribuzione congiunta e marinali è illustrato in fig. 5. Nel caso che stiamo

considerando di variabili indipendenti è

$$f_{M1}(x_1) = f_{C1}(x_1, x_2^*) = f_1(x_1), \quad f_{M2}(x_2) = f_{C2}(x_2, x_1^*) = f_2(x_2).$$

Le intersezioni con piani orizzontali, paralleli al piano  $x_1, x_2$ , corrispondono a valori costanti della funzione di distribuzione  $f(x_1, x_2)$ , cioè a valori costanti  $c$  di  $Q^2$ . Le curve corrispondenti sono delle ellissi centrate in  $(\mu_1, \mu_2)$  di equazione

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = c.$$

Per  $c = 1$  l'ellisse ha semiassi  $\sigma_1$  e  $\sigma_2$ . La probabilità che una coppia di valori  $x_1, x_2$  cada nel rettangolo circoscritto all'ellisse è  $P_{r1} = 0.68 \times 0.68$ . La probabilità che la coppia cada all'interno dell'ellisse è  $P_{e1} = 0.39$ . Infatti,  $Q^2$  è la somma dei quadrati di due variabili con distribuzione normale standard e quindi, come già detto, è una variabile con funzione di distribuzione di  $\chi^2$  a 2 gradi di libertà. La probabilità che sia  $Q^2 < 1$  è la probabilità di avere  $\chi_2^2 < 1$ , pari, appunto, a 0.39, come si ricava facilmente integrando la distribuzione. Per  $c = 4$  l'ellisse ha semiassi  $2\sigma_1$  e  $2\sigma_2$ , e la probabilità che una coppia cada al suo interno è la probabilità che sia  $\chi_2^2 < 4$  cioè 0.86.

Nel caso di  $n$  variabili casuali indipendenti  $x_i$  con funzioni di distribuzione di Gauss  $N(\mu_i, \sigma_i^2)$ , la funzione di distribuzione congiunta è ancora data dal prodotto delle funzioni di distribuzione:

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma_1 \dots \sigma_n} e^{-Q^2/2} \quad \text{con} \quad Q^2 = \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2}.$$

Usando la matrice delle covarianze

$$V = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

che è diagonale trattandosi di variabili indipendenti, e introducendo i vettori  $\vec{x}$  e  $\vec{\mu}$  con componenti  $x_i$  e  $\mu_i$ , la funzione di distribuzione congiunta si può scrivere come

$$f(\vec{x}) = \frac{1}{(2\pi)^{n/2} |\det V|^{1/2}} e^{-Q^2/2} \quad \text{con} \quad Q^2 = (\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu})$$

Questa è l'espressione generale della funzione di distribuzione multinormale, ed è la stessa anche nel caso in cui i coefficienti di correlazione siano diversi da zero, e quindi la matrice

delle covarianze sia

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{21}\sigma_2\sigma_1 & \cdots & \rho_{n1}\sigma_n\sigma_1 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{n2}\sigma_n\sigma_2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix}$$

L'espressione esplicita della funzione di distribuzione multinormale si può ottenere facilmente nel caso di due sole variabili  $x_1$  e  $x_2$ . In questo caso la matrice delle covarianze diventa

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

il cui determinante è  $\det V = \sigma_1^2\sigma_2^2(1 - \rho_{12}^2)$ . Si ha quindi

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1 - \rho_{12}^2)^{1/2}} e^{-Q^2/2}$$

con

$$Q^2 = \frac{1}{1 - \rho_{12}^2} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho_{12} \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right].$$

Per  $Q^2 = c$  costante, si ottiene ancora un'ellisse, ancora con centro  $(\mu_1, \mu_2)$  ma questa volta non riferita agli assi principali; l'angolo di rotazione  $\phi$  può essere maggiore o minore di  $\pi/2$ , in accordo con il segno del coefficiente di correlazione.

Per  $Q^2 = 1$  si ottiene l'ellisse rappresentata schematicamente in fig. 6. Considerando i punti di intersezione dell'ellisse con rette parallele agli assi  $x_1$  e  $x_2$  si può vedere che:

- la retta  $x_1 = \mu_1$  interseca l'ellisse nei punti di coordinate  $(\mu_2 \pm (1 - \rho_{12}^2)^{1/2} \sigma_2)$ ;
- la retta  $x_2 = \mu_2$  interseca l'ellisse nei punti di coordinate  $(\mu_1 \pm (1 - \rho_{12}^2)^{1/2} \sigma_1)$ ;
- le equazioni delle rette tangenti parallele all'asse  $x_2$  sono  $x_1 = \mu_1 \pm \sigma_1$  e i punti di tangenza hanno coordinate  $(\mu_1 \pm \sigma_1, \mu_2 \pm \rho_{12}\sigma_2)$ ;
- le equazioni delle rette tangenti parallele all'asse  $x_1$  sono  $x_2 = \mu_2 \pm \sigma_2$  e i punti di tangenza hanno coordinate  $(\mu_1 \pm \rho_{12}\sigma_1, \mu_2 \pm \sigma_2)$ .

Da queste relazioni si possono determinare deviazioni standard e coefficiente di correlazione, una volta nota l'ellisse. Notare che per  $\rho_{12} = 0$  l'ellisse è riferita agli assi principali, e che per  $\rho_{12} = 1$  è un segmento.

L'ellisse  $Q^2 = 4$  ha stessa inclinazione ma assi di lunghezza doppia rispetto all'ellisse  $Q^2 = 1$  e corrisponde quindi a due deviazioni standard.

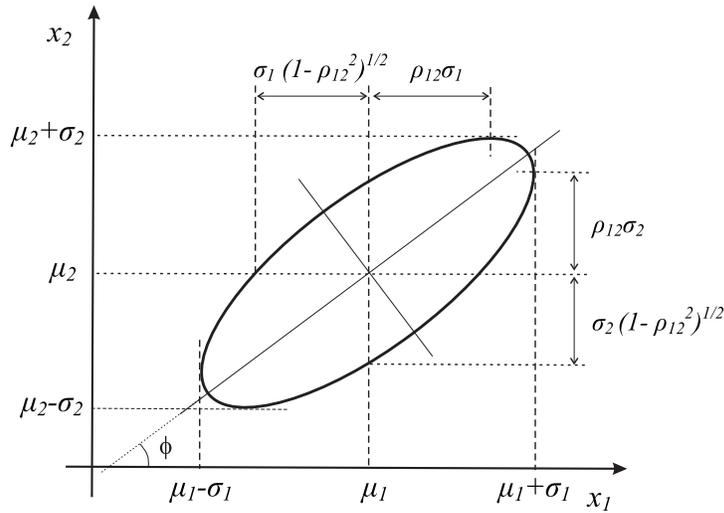


Figura 6: Ellisse corrispondente a  $Q^2 = 1$  (da PDG).

Le funzioni di distribuzione marginali sono, anche in questo caso, le distribuzioni di Gauss delle singole variabili:

$$f_{M1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}}$$

e l'espressione è analoga per  $f_{M2}(x_2)$ .

\* dimostrazione

la forma quadratica  $Q^2$  può essere essere riscritta come

$$Q^2 = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu'_2)^2}{\sigma_2'^2}$$

dove  $\mu'_2 = \mu_2 + \rho_{12}(x_1 - \mu_1)\sigma_2/\sigma_1$  e  $\sigma_2'^2 = \sigma_2^2(1 - \rho_{12}^2)$ . La funzione di distribuzione congiunta può quindi essere scritta come il prodotto di due funzioni di distribuzione normali:

$$f(x_1, x_2) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2'} e^{-\frac{(x_2 - \mu'_2)^2}{2\sigma_2'^2}}$$

da cui, integrando su  $x_2$ , si ottiene  $f_{M1}(x_1)$ .

Si può poi ricavare che la funzione di distribuzione condizionata di  $x_1$  (e per  $x_2$ , con  $x_2 = x_2^*$  per un valore fissato  $x_2^*$  di  $x_2$  è

$$f_{C1}(x_1, x_2^*) = \frac{f(x_1, x_2^*)}{f_{M2}(x_2^*)} = \frac{1}{\sqrt{2\pi}\sigma_1'} e^{-\frac{(x_1 - \mu_1')^2}{2\sigma_1'^2}}$$

con

$$\sigma_1' = \sigma_1(1 - \rho_{12})^{1/2}, \quad \mu_1' = \mu_1 + \rho_{12}\frac{\sigma_1}{\sigma_2}(x_2^* - \mu_2).$$

\* dimostrazione

basta scrivere esplicitamente la funzione di distribuzione condizionata di  $x_1$  usando la funzione di distribuzione marginale di  $x_2$ :

$$f_{C1}(x_1, x_2^*) = \frac{f(x_1, x_2^*)}{f_{M2}(x_2^*)} = \frac{1}{\sqrt{2\pi}\sigma_1(1-\rho_{12})^{1/2}} e^{-Q^{*2}/2}$$

dove

$$\begin{aligned} Q^{*2} &= \frac{1}{1-\rho_{12}^2} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho_{12} \frac{x_1 - \mu_1}{\sigma_1} \frac{x_2^* - \mu_2}{\sigma_2} + \frac{(x_2^* - \mu_2)^2}{\sigma_2^2} \right] - \frac{(x_2^* - \mu_2)^2}{\sigma_2^2} \\ &= \frac{(x_1 - \mu_1')^2}{\sigma_1'^2} \end{aligned}$$

La varianza è quindi costante e ridotta del fattore  $(1-\rho_{12})$ , mentre il valore di aspettazione di  $x_1$  dipende da  $x_2^*$ . Le rette

$$\begin{aligned} x_1 &= \mu_1 + \rho_{12} \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2) \\ x_2 &= \mu_2 + \rho_{12} \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1) \end{aligned}$$

sono i luoghi dei punti di massimo delle funzioni di distribuzione condizionate di  $x_1$  e  $x_2$ , e sono dette rette di regressione. Sono le rette che congiungono le due coppie di punti di tangenza di coordinate  $(\mu_1 \pm \rho_{12}\sigma_1, \mu_2 \pm \sigma_2)$  e  $(\mu_1 \pm \sigma_1, \mu_2 \pm \rho_{12}\sigma_2)$ .

L'importanza della distribuzione multinormale sta nel fatto che è molto diffusa. Spesso si utilizzano campioni con distribuzione multinormale e, come verrà dimostrato e si vedrà nella parte pratica, in certi casi le stime dei parametri hanno distribuzione multinormale.

## 1.4 Funzioni di una variabile casuale

La trattazione delle funzioni di variabili casuali in teoria della probabilità è un argomento molto importante per le sue applicazioni in inferenza statistica. Come vedremo, gli stimatori e le funzioni di test dei test d'ipotesi sono infatti funzioni di un campione casuale, e quindi funzioni di più variabile casuale.

In questo capitolo vengono riassunte le considerazioni sulle funzioni di una variabile casuale. Il prossimo capitolo è dedicato alle funzioni di più variabili casuali e, dopo le considerazioni di carattere generale, verranno considerate alcune funzioni di variabili casuali particolarmente interessanti, usate nella restante parte del corso.

In generale, se  $y = y(x)$  è una funzione della variabile casuale  $x$  con funzione di distribuzione  $f(x)$ , anche  $y$  è una variabile casuale con una certa funzione di distribuzione  $g(y)$ . Per il significato di funzione di distribuzione, nel caso di corrispondenza biunivoca tra  $x$  e  $y$ , deve essere

$$g(y)dy = f(x)dx$$

e quindi

$$g(y) = f[x(y)] \left| \frac{dx(y)}{dy} \right|.$$

(se, ovviamente, se la funzione  $y = y(x)$  si può invertire).

Se la funzione non è a un solo valore, cioè se esistono  $m$  valori  $x_j$  tali che  $y(x_j) = y$ , la funzione di distribuzione di  $y$  è

$$g(y) = \sum_{j=1}^m f[x_j(y)] \left| \frac{dx_j(y)}{dy} \right|.$$

Usando queste relazioni, si ottiene facilmente che:

- se  $x$  ha funzione di distribuzione  $N(\mu_x, \sigma_x^2)$ ,  $y = (x - \mu_x)/\sigma_x$  ha funzione di distribuzione  $N(0, 1)$ ;
- se  $x$  ha funzione di distribuzione  $N(0, 1)$ ,  $y = \sigma(x + \mu)$  ha funzione di distribuzione  $N(\mu, \sigma^2)$ ;
- se  $x$  ha funzione di distribuzione  $N(0, 1)$ ,  $y = x^2$  ha funzione di distribuzione di  $\chi^2$  a un grado di libertà.
- ....

A volte, comunque, non è necessario calcolare la funzione di distribuzione di  $y$  ma è sufficiente calcolarne valore di aspettazione e varianza con le formule approssimate

$$\mu_y = E[y] \simeq y(\mu_x) + \frac{1}{2} \left[ \frac{d^2 y}{dx^2} \right]_{x=\mu_x} \sigma_x^2 \simeq y(\mu_x),$$

$$\sigma_y^2 = \text{var}(y) \simeq \left( \left[ \frac{dy}{dx} \right]_{x=\mu_x} \right)^2 \sigma_x^2$$

valide qualunque sia la funzione di distribuzione di  $x$ . Sono le formule già usate per le misure indirette e si ottengono usando lo sviluppo in serie di  $y(x)$  in un intorno di  $x = \mu_x$ .

## 1.5 Funzioni di più variabili casuali

Generalizzando quanto visto per una variabile casuale, nel caso di funzione di più variabili casuali,  $y = y(x_1, x_2, \dots, x_n)$  con  $x_i$   $n$  variabili casuali di cui sono noti valori di aspettazione e varianze, si può ancora utilizzare lo sviluppo in serie di  $y(x_1, x_2, \dots, x_n)$  per scrivere le formule approssimate per valore aspettazione e varianza di  $y$ :

$$\mu_y = E[y] \simeq y(\mu_{x_1}, \mu_{x_2}, \dots, \mu_{x_n}),$$

$$\begin{aligned} \sigma_y^2 &\simeq \sum_{k=1}^n \left( \left[ \frac{\partial y}{\partial x_k} \right]_{\vec{x}=\vec{\mu}_x} \right)^2 \sigma_{x_k}^2 + \\ &+ \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[ \frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}_x} \left[ \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}_x} \text{cov}(x_i, x_j). \end{aligned}$$

in cui, per variabili  $x_i$  correlate, i termini contenenti le covarianze non possono essere trascurati. Questi possono essere positivi o negativi, a seconda del segno della covarianza e delle derivate parziali, e quindi possono aumentare o diminuire l'incertezza su  $y$  (vedi esempi).

Notare che due funzioni  $y_1$  e  $y_2$  delle stesse variabili casuali  $x_1, x_2, \dots, x_n$  hanno covarianza diversa da zero anche se le  $x_i$  sono indipendenti. Usando  $\text{cov}(y_1, y_2) = E[y_1 y_2] - E[y_1]E[y_2]$  e lo sviluppo in serie di  $y_1$  e  $y_2$  per calcolare i valori di aspettazione, si ottiene

$$\text{cov}(y_1, y_2) \simeq \sum_{i=1}^n \left[ \frac{\partial y_1}{\partial x_i} \right]_{\vec{x}=\vec{\mu}_x} \left[ \frac{\partial y_2}{\partial x_i} \right]_{\vec{x}=\vec{\mu}_x} \sigma_{x_i}^2.$$

Usando questa formula si può, ad esempio, calcolare la correlazione fra le stime dei parametri di una retta ottenute con il metodo dei minimi quadrati.

Per ottenere la funzione di distribuzione di  $y = y(x_1, x_2, \dots, x_n)$  con  $x_i$  variabili casuali con funzione di distribuzione congiunta  $f(x_1, x_2, \dots, x_n)$ , nel caso generale è necessario

introdurre  $n$  variabili casuali  $y_i$  funzioni delle  $n$  variabili casuali  $x_i$ :

$$\begin{aligned} y_1 &= y_1(x_1, x_2, \dots, x_n) \\ y_2 &= y_2(x_1, x_2, \dots, x_n) \\ &\dots \\ y_n &= y_n(x_1, x_2, \dots, x_n) \end{aligned}$$

di cui, ad esempio, sarà  $y_1 = y$ . Se esistono le trasformazioni inverse  $x_k = x_k(y_1, y_2, \dots, y_n)$ , la funzione di distribuzione congiunta di  $y_1, y_2, \dots, y_n$  è

$$g(y_1, y_2, \dots, y_n) = f[x_1(y_1, y_2, \dots, y_n), x_2(y_1, y_2, \dots, y_n), \dots, x_n(y_1, y_2, \dots, y_n)] \cdot |J|$$

dove

$$J = \left| \frac{\partial(x_1, x_2, \dots, x_n)}{\partial(y_1, y_2, \dots, y_n)} \right|$$

$$= \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \dots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

è il determinante dello Jacobiano della trasformazione. Naturalmente l'espressione diventa più complessa nel caso di funzioni a più valori. Una volta calcolata la funzione di distribuzione congiunta  $g(y_1, y_2, \dots, y_n)$ , integrandola su tutte le variabili  $y_j \neq y_1$ , si ottiene la funzione di distribuzione marginale di  $y_1$ .

Esempi:  $y = x_1 \pm x_2$ ,  $y = (-2 \ln x_1)^{1/2} \cos 2\pi x_2$ , con  $x_1, x_2$  variabili casuali con distribuzione uniforme in  $(0,1)$

Non sempre questa procedura è di facile applicazione, ma, come vedremo nel seguito, in alcuni casi importanti è possibile ottenere la funzione di distribuzione di una funzione di variabili casuali con metodi più semplici, come l'uso delle funzioni generatrici dei momenti.

La parte rimanente di questo capitolo, che conclude la trattazione della teoria della probabilità, è dedicata ad alcune funzioni di variabili casuali particolarmente importanti nell'analisi dei dati sperimentali e per l'inferenza statistica.

### 1.5.1 Somma di quadrati di variabili casuali

Nel caso in cui le variabili casuali indipendenti  $X_i$  abbiano distribuzione normale standard  $N(0, 1)$ , allora la somma dei loro quadrati

$$Z = \sum_{i=1}^n X_i^2. \quad (1)$$

ha una funzione di distribuzione di  $\chi^2$  a  $\nu = n$  gradi di libertà.

\* Infatti, come visto nel paragrafo 1.4, il quadrato di una variabile casuale  $X$  con distribuzione  $N(0, 1)$  è una variabile casuale con funzione di distribuzione di  $\chi_1^2$  la cui funzione generatrice dei momenti algebrici è  $M_X^*(t) = E[e^{tX^2}] = (1 - 2t)^{-1/2}$ . La funzione generatrice dei momenti algebrici della funzione di distribuzione di  $Z$  è quindi

$$M_Z^*(t) = E[e^{t\sum_{i=1}^n X_i^2}] = \prod_i E[e^{tX_i^2}] = (1 - 2t)^{-n/2} \quad (2)$$

che è la funzione generatrice dei momenti algebrici della funzione di distribuzione di  $\chi_n^2$ .

Ne segue che date  $n$  variabili  $X_i$  con funzione di distribuzione  $N(\mu_i, \sigma_i^2)$ , le variabili  $Y_i = (X_i - \mu_i)/\sigma_i$  con funzione di distribuzione  $N(0, 1)$  e quindi la somma dei loro quadrati

$$Z = \sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2} \quad (3)$$

ha funzione di distribuzione di  $\chi^2$  a  $\nu = n$  gradi di libertà.

Questa è la base per un test d'ipotesi molto diffuso, il test di  $\chi^2$ .

Inoltre a questo punto è chiaro perchè la funzione  $Q^2$  introdotta nel paragrafo 1.3.6 ha funzione di distribuzione di  $\chi^2$  a  $n$  gradi di libertà, dove  $n$  è il numero di variabili  $x_i$  con funzione di distribuzione di Gauss, almeno per variabili indipendenti.

Se le variabili sono correlate

$$Z = (\vec{X} - \vec{\mu})^T V^{-1} (\vec{X} - \vec{\mu}) \quad (4)$$

ha funzione di distribuzione di  $\chi^2$  a  $n$  gradi di libertà. Infatti è sempre possibile (se le variabili non sono ridondanti e quindi  $V$  non è singolare) trovare una trasformazione ortogonale che permetta di passare dalle  $n$  variabili correlate  $X_i$  a  $n$  variabili indipendenti  $Y_i$  per le quali la matrice delle covarianze è diagonale. La dimostrazione è meno immediata ed è stata fatta esplicitamente solo nel caso  $n = 2$ .

Le applicazioni sono molte altre. Ad esempio, se consideriamo un vettore  $\vec{r} = (r_1, r_2, r_3)$  le cui componenti abbiano distribuzione  $N(0, \sigma^2)$ ,  $r^2 = r_1^2 + r_2^2 + r_3^2$  ha una distribuzione  $\sigma^2 \chi_3^2$ . Il modulo di  $\vec{r}$  ha quindi una funzione di distribuzione che si può facilmente calcolare. Questo permette di ottenere, come già visto, la distribuzione di Maxwell-Boltzmann delle velocità molecolari di un gas, e da questa la velocità media, ecc.

### 1.5.2 Somma di variabili casuali

Consideriamo la variabile casuale  $y$  definita come somma di  $n$  variabili casuali indipendenti  $X_i$ . Usando il fatto che la funzione generatrice dei momenti di una somma di variabili casuali è il prodotto delle funzioni generatrici dei momenti delle singole variabili, si può vedere che:

- se le variabili  $X_i$  hanno tutte funzione di distribuzione  $N(\mu, \sigma^2)$ ,
- se le variabili  $X_i$  hanno funzione di distribuzione  $N(\mu_i, \sigma_i^2)$ , la funzione di distribuzione di  $Y$  è  $N(\sum_i \mu_i, \sum_i \sigma_i^2)$
- se le variabili  $X_i$  hanno funzioni di distribuzione  $\chi_{\nu_i}^2$ ,  $Y$  ha una funzione di distribuzione di  $\chi^2$  con  $\nu = \sum_{i=1}^r \nu_i$  gradi di libertà. Questa proprietà (proprietà additiva del  $\chi^2$ ) è molto usata, ad esempio quando si vogliono combinare misure provenienti da diversi esperimenti.

\* dimostrazioni usando le funzioni generatrici dei momenti

### 1.5.3 Media aritmetica

La media aritmetica di un campione di dimensione  $n$ , come definito nel paragrafo precedente,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (5)$$

gode di proprietà molto importanti che ne giustificano l'ampio uso.

La media è centrata attorno al valore di aspettazione di  $X$ , infatti

$$E[\bar{X}] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu \quad (6)$$

e, se le  $X_i$  sono indipendenti, come ipotizziamo, è

$$var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n var(X_i) = \frac{\sigma^2}{n}, \quad (7)$$

indipendentemente dalla funzione di distribuzione  $f(X)$ .

Altre proprietà della media sono molto importanti.. In particolare la Legge dei Grandi Numeri afferma che che la media converge in probabilità a  $\mu$ , cioè che

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0 \quad (8)$$

dove  $\epsilon$  è un numero positivo piccolo a piacere.

\* dimostrazione per  $\sigma^2$  finita

La Legge dei Grandi Numeri è valida per qualsiasi  $f(X)$ , anche nel caso in cui la varianza non sia finita, come per la distribuzione di Cauchy.

Se  $f(X)$  è  $N(\mu, \sigma^2)$ ,  $\bar{X}$  ha distribuzione  $N(\mu, \sigma^2/n)$ . Non solo, ma vale il Teorema del Limite Centrale: per  $n \rightarrow \infty$ ,  $\bar{X}$  ha distribuzione  $N(\mu, \sigma^2/n)$  qualunque sia  $f(X)$  purchè  $\sigma^2$  sia finita.

\* dimostrazione

In pratica, basta che sia  $n$  grande perchè questo accada. Da qui l'importanza della media, e la grande diffusione della distribuzione di Gauss. In particolare si può capire come l'ipotesi di molti effetti che intervengono nelle operazioni di misure porti alla distribuzione degli errori accidentali.

Infine, nel caso in cui le funzioni di distribuzione delle  $X_i$  siano diverse, e siano diversi valori di aspettazione e varianze, si può dimostrare che la media ha ancora distribuzione normale con

$$E[\bar{X}] = \frac{1}{n} \sum_{i=1}^n \mu_i \text{ e } \text{var}(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2. \quad (9)$$

#### 1.5.4 Somme di quadrati degli scarti

Da quanto visto nel paragrafo precedente, con riferimento al campione di dimensione  $n$ , si ottiene facilmente che la somma dei quadrati degli scarti dal valore di aspettazione

$$Z = \sum_{i=1}^n (X_i - \mu)^2 \quad (10)$$

ha, a parte un fattore di scala  $\sigma^2$ , una funzione di distribuzione di  $\chi^2$  con  $\nu = n$  gradi di libertà. Il suo valore di aspettazione e la sua varianza sono quindi

$$E[Z] = \sigma^2 E[\chi_n^2] = n\sigma^2, \quad \sigma_Z^2 = (\sigma^2)^2 \text{var}(\chi_n^2) = 2n(\sigma^2)^2. \quad (11)$$

Nel caso in cui non si conosca  $\mu$ , si può introdurre la variabile

$$Z = \sum_{i=1}^n (X_i - \bar{X})^2 \quad (12)$$

cioè la somma dei quadrati degli scarti dalle media. In questo caso però gli addendi non sono tutti indipendenti perchè interviene la media, che è una relazione tra le variabili  $X_i$ . Con un'opportuna trasformazione di variabili, però,  $z$  può essere scritta come la somma dei quadrati di  $n - 1$  variabili indipendenti tutte con valore di aspettazione 0 e varianza

$\sigma^2$ . Essendo, in generale,  $\sigma_x^2 = E[x^2] - (E[x])^2$ , è  $E[Z] = (n-1)\sigma^2$ . Ne consegue che la varianza del campione

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} (\overline{X^2} - \bar{X}^2) \quad (13)$$

ha valore di aspettazione  $E[s^2] = \sigma^2$ , ed è quindi una stima centrata della varianza.

\* dimostrazione

Inoltre, se  $f(X)$  è una distribuzione di Gauss, allora  $s^2$  ha una funzione di distribuzione di  $\chi^2$  a  $\nu = n-1$  gradi di libertà, a parte un fattore  $\sigma^2/(n-1)$ . Questo permette di calcolare subito varianza e deviazione standard di  $s^2$  e deviazione standard di  $\sqrt{s^2}$ .

\* ... vedi appunti ...

### Covarianza e coefficiente di correlazione

Formule analoghe a eq. (13) permettono di costruire stime centrate della covarianza di due variabili casuali. Date due variabili  $X, Y$  con valori di aspettazione  $\mu_x, \mu_y$ , la loro covarianza è

$$\text{cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]. \quad (14)$$

Generalmente si ha a disposizione un campione costituito da  $n$  coppie di valori  $(x_i, y_i)$ , e  $\mu_x, \mu_y$  non sono noti, ma vengono stimati usando le medie aritmetiche  $\bar{X}, \bar{Y}$ . Si può dimostrare che

$$v_{x,y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{n-1} (\overline{XY} - \bar{X} \cdot \bar{Y}) \quad (15)$$

ha come valore di aspettazione  $\text{cov}(X, Y)$ , ed è quindi uno stimatore centrato della covarianza.

Per stimare il coefficiente di correlazione  $\rho = \text{cov}(X, Y)/(\sigma_x \sigma_y)$  si può quindi pensare di usare le stime per covarianze e varianze, cioè

$$r = \frac{v_{x,y}}{\sqrt{s_X^2 s_Y^2}} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\sqrt{(\overline{X^2} - \bar{X}^2)(\overline{Y^2} - \bar{Y}^2)}}. \quad (16)$$

Benchè normalmente si usi effettivamente  $r$ , è bene ricordare che  $r$  non è una stima centrata, nonostante lo siano  $v_{x,y}$  e  $s_X^2, s_Y^2$  (lo è solo asintoticamente), ed ha varianza piuttosto grande. Infatti è

$$E[r] = \rho - \frac{\rho(1-\rho^2)}{2n} + \dots, \quad \sigma_r^2 = \frac{1}{n}(1-\rho^2)^2 + \dots$$

### Esercizi

1. Ricavare le funzioni generatrici dei momenti della funzione di distribuzione binormale nel caso di coefficiente di correlazione diverso da zero.
2. Dimostrare che, se  $x_1$  e  $x_2$  sono variabili casuali indipendenti con funzione di distribuzione uniforme tra 0 e 1,

$$z_1 = \sqrt{-2 \ln x_1} \cos(2\pi x_2), \quad z_2 = \sqrt{-2 \ln x_1} \sin(2\pi x_2)$$

sono variabili casuali indipendenti con funzione di distribuzione binormale e funzione di distribuzione marginale  $N(0; 1)$

3. Calcolare la deviazione standard  $\sigma_{s^2}$  della varianza  $s^2$  di un campione di dimensione  $n$  e funzione di distribuzione normale  $N(\mu; \sigma^2)$  per  $n = 2, 5, 10, 50$ .  
Calcolare la probabilità  $P_k$  che  $s^2$  appartenga agli intervalli  $(\sigma^2 - k\sigma_{s^2}, \sigma^2 + k\sigma_{s^2})$  con  $k = 1, 2, 3$ .

## 2 Inferenza statistica

In statistica si è interessati ad utilizzare un insieme di dati per ottenere informazioni su un modello probabilistico, cioè investigare la validità del modello e/o determinare il valore dei parametri che vi compaiono. Con inferenza statistica (o statistica inferenziale) si indica il procedimento per cui si inducono le caratteristiche di una popolazione dall'osservazione di una parte di essa, il campione casuale o aleatorio, che si assume rappresentativa della popolazione. Il primo punto è quindi la selezione di un modello statistico per il processo che genera i dati. Un esempio è costituito da misure ripetute nelle stesse condizioni di una grandezza fisica, con valore vero ben definito, con errori accidentali dominanti. In questo caso, il modello statistico, ben giustificato in certe ipotesi sulla natura degli errori accidentali, è che la distribuzione delle misure sia gaussiana con valore di aspettazione il valore vero della grandezza fisica. Avendo a disposizione una serie di misure (il campione) si vuole ottenere la stima del valore vero della grandezza fisica (come noto, la media aritmetica dei valori misurati) e valutare la validità del modello ipotizzato (errori accidentali dominanti, funzione di distribuzione di Gauss per le misure).

Ci sono due approcci diversi all'inferenza statistica, l'inferenza frequentista e l'inferenza bayesiana, che si differenziano essenzialmente per l'interpretazione di probabilità e di conseguenza per l'utilizzo e l'importanza che si dà al teorema di Bayes.

Nella statistica frequentista, l'unica trattata nel corso, la probabilità è interpretata come la frequenza di un risultato in un esperimento ripetibile. I punti più importanti, trattati brevemente nei prossimi capitoli, sono:

- stima dei parametri;
- intervalli di confidenza, costruiti per includere, con una certa probabilità, i valori veri dei parametri;
- test statistici, tra cui il test d'ipotesi.

In statistica bayesiana, l'interpretazione di probabilità è più generale, include la probabilità soggettiva e permette di usare informazioni aggiuntive, in genere soggettive. In particolare si introduce anche la funzione densità di probabilità del parametro. In molti casi, i risultati numerici dei due approcci sono gli stessi, ma in alcuni casi i risultati possono ovviamente essere molto diversi.

## 3 Stima dei parametri

### 3.1 Considerazioni generali

La stima dei parametri è un altro aspetto di grande importanza dell'inferenza statistica. Da un numero limitato di osservazioni, che si assume costituiscano un "campione casuale", si vogliono ottenere informazioni quantitative sull'intera popolazione di cui il campione fa parte. In particolare, la forma matematica della distribuzione che descrive una certa popolazione può essere ben definita (nota o ipotizzata) ma possono non essere noti i valori numerici dei parametri che compaiono nell'espressione della distribuzione di probabilità. Il problema è proprio determinare questi valori numerici a partire dal campione. Dalle misure disponibili dovrebbero quindi essere estratte più informazioni possibili sui parametri e comunque numeri che rappresentino i valori dei parametri, cioè effettuare la "stima dei parametri".

Esistono diversi metodi che possono essere applicati per stimare i parametri che intervengono nei fenomeni che si stanno studiando. Ne vedremo due (i metodi del Maximum Likelihood, o massima verosimiglianza, e dei Least Squares, o minimi quadrati) che forniscono "stime puntuali" dei parametri, cioè valori numerici, e le incertezza sui valori stessi.

Prima di considerare questi metodi, è necessario chiarire alcune definizioni e, come verrà fatto nel prossimo paragrafo, considerare le caratteristiche generali delle stime.

#### Definizioni

##### **Stimatore** ("estimator")

è una funzione  $t$  delle misure (o un metodo o una prescrizione sull'utilizzo del campione) usata per trovare il valore del parametro (o dei parametri) incogniti:  $\hat{\theta} = t(X_1, X_2, \dots, X_n)$ , dove  $X_1, X_2, \dots, X_n$  è il campione casuale. È quindi una statistica, una funzione del campione. Per uno stesso parametro diversi metodi possono suggerire l'uso di stimatori diversi (ad esempio, per stimare il valore di aspettazione di  $X$  si potrebbero usare sia la media che la mediana del campione), che però non avranno in generale le stesse caratteristiche.

##### **Stima** ("estimate")

è il valore numerico che lo stimatore assume sullo specifico campione:  $\hat{\theta}_0 = t(x_1, x_2, \dots, x_n)$ . Il termine "stimatore" è in genere poco usato, e anche in queste note si userà spesso "stima" al suo posto. Dal contesto dovrebbe comunque essere sempre chiaro se si parla della funzione o del valore numerico.

##### **Likelihood**

Supponiamo che la variabile casuale  $X$  abbia una funzione di distribuzione  $f(X, \vec{\theta})$ , dove  $\vec{\theta}$  è il vettore con componenti i parametri (ad es., per la funzione di distribuzione di Gauss  $\vec{\theta} = (\mu, \sigma^2)$  nella notazione usuale). Se abbiamo a disposizione un campione casuale di

dimensione  $n$  costituito da  $n$  misure indipendenti, la funzione di likelihood è

$$f(X_1, X_2, \dots, X_n, \vec{\theta}) = \prod_{i=1}^n f(X_i, \vec{\theta}). \quad (17)$$

La funzione di likelihood è quindi la densità di probabilità congiunta delle variabili  $X_i$  se  $\vec{\theta}$  è noto, quindi un numero se calcolata sullo specifico campione. Può anche essere pensata come la densità di probabilità condizionata per le  $X_i$  dato  $\vec{\theta}$ . Infine, la funzione di likelihood sullo specifico campione  $x_1, x_2, \dots, x_n$  è una funzione dei soli parametri e può essere pensata come la funzione densità di probabilità dei parametri.

### 3.2 Caratteristiche degli stimatori

Gli stimatori sono funzioni di variabili casuali (il campione), e quindi variabili casuali che assumono valori puntuali su un certo campione. Le loro distribuzioni, che ne riflettono la qualità, sono quindi molto importanti. Ad esempio, nella misura di una grandezza fisica con un certo “valore vero” in presenza di errori accidentali non trascurabili, è ben noto che una stima del valore vero è data dalla media aritmetica dei valori ottenuti ripetendo la misura nelle stesse condizioni. Il valore numerico in se, però, se non sapessimo qual’è la funzione di distribuzione della media, non sarebbe sufficiente.

Nel seguito sono descritte alcune delle caratteristiche probabilistiche che un buon stimatore dovrebbe possedere. Verrà considerato il caso di un parametro, e lo stimatore, funzione del campione  $X_1, X_2, \dots, X_n$ , verrà indicato con  $t$ .

#### Consistente

Lo stimatore  $t$  è “consistente” se converge in probabilità al valore vero del parametro. Ciò significa che, se  $\hat{\theta}_n = t(x_1, x_2, \dots, x_n)$  è il valore che lo stimatore assume su uno specifico campione e  $\theta_0$  il valore vero parametro, è

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta_0| \leq \epsilon) = 1. \quad (18)$$

Detto in altro modo, esiste un valore  $N$  tale che, per  $n > N$ , è

$$P(|\hat{\theta}_n - \theta_0| \leq \epsilon) \geq \eta, \quad (19)$$

con  $\epsilon$  e  $\eta$  numeri positivi arbitrari.

\* esempio: media aritmetica di misure ripetute di una grandezza fisica

È chiaro che un “buon” stimatore deve essere consistente.

#### Centrato

Uno stimatore si dice “centrato” (o unbiased, senza bias) se il suo valore di aspettazione è il valore vero del parametro, cioè se  $E[t] = \theta_0$ .

Se invece è  $E[t] = \theta_0 + b$ , lo stimatore non è centrato. Nel caso il cui sia

$$\lim_{n \rightarrow \infty} b = 0, \quad (20)$$

lo stimatore si dice “asintoticamente centrato”.

### Varianza minima e efficienza

Essendo la varianza una buona misura della concentrazione delle stime, tra i possibili stimatori centrati e consistenti si sceglie quello a varianza minore. Si può dimostrare ad esempio che la media del campione ha varianza minore della mediana, ed è per questo che in genere viene usata la prima.

La scelta è facilitata dal fatto che c'è un limite inferiore per la varianza delle stime dei parametri che si possono ottenere da un certo campione casuale  $\vec{X}$ . Si può infatti dimostrare che se  $\tau(\theta)$  è una funzione del parametro  $\theta$  (con eventualmente  $\tau(\theta) = \theta$ ) di cui la statistica  $t(\vec{X})$  è uno stimatore, vale la disuguaglianza di Cramer-Rao (o di Cramer-Rao-Frechèt):

$$\sigma_t^2 \geq \frac{\left(\frac{d\tau}{d\theta} + \frac{db}{d\theta}\right)^2}{E\left[-\frac{\partial^2 \ln L}{\partial \theta^2}\right]} \quad (21)$$

dove  $b$  è l'eventuale bias.

\* dimostrazione:

Nel caso generale  $t$  non è centrato e si ha  $E[t] = \tau(\theta) + b(\theta) = \int_{\vec{\Omega}_x} d\vec{X} t L(\vec{X}, \theta)$ . Qui  $\theta$  indica lo specifico valore del parametro che compare nella funzione di distribuzione congiunta del campione  $L(\vec{X}, \theta)$ , utilizzata per calcolare i valori di aspettazione. Vogliamo calcolare la varianza di  $t$ , cioè  $\sigma_t^2 = E[(t - \tau)^2]$ . Si ha:

$$\frac{dE[t]}{d\theta} = \int_{\vec{\Omega}_x} d\vec{X} t \frac{\partial L}{\partial \theta} = \int_{\vec{\Omega}_x} d\vec{X} t L \frac{\partial \ln L}{\partial \theta} = E\left[t \frac{\partial \ln L}{\partial \theta}\right]$$

ed è anche

$$\frac{dE[t]}{d\theta} = \frac{d\tau}{d\theta} + \frac{db}{d\theta}, \quad E\left[t \frac{\partial \ln L}{\partial \theta}\right] = \frac{d\tau}{d\theta} + \frac{db}{d\theta}.$$

Per la condizione di normalizzazione di  $L$  è inoltre  $\int_{\vec{\Omega}_x} d\vec{X} L = 1$  e quindi, derivando rispetto a  $\theta$ :

$$\int_{\vec{\Omega}_x} d\vec{X} \frac{\partial L}{\partial \theta} = 0, \quad \int_{\vec{\Omega}_x} d\vec{X} L \frac{\partial \ln L}{\partial \theta} = 0;$$

moltiplicando per  $\tau$  si ottiene

$$\int_{\vec{\Omega}_x} d\vec{X} \tau L \frac{\partial \ln L}{\partial \theta} = E\left[\tau \frac{\partial \ln L}{\partial \theta}\right] = 0.$$

Sottraendo le due espressioni trovate si ottiene

$$E\left[(t - \tau) \frac{\partial \ln L}{\partial \theta}\right] = \frac{d\tau}{d\theta} + \frac{db}{d\theta}.$$

Ricordando la disuguaglianza di Schwarz,  $(E[xy])^2 \leq E[x^2] \cdot E[y^2]$  dove l'uguaglianza vale se  $x = ky$ , si ha

$$E[(t - \tau)^2] \cdot E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right] \geq \left(\frac{d\tau}{d\theta} + \frac{db}{d\theta}\right)^2$$

e quindi

$$\sigma_t^2 \geq \frac{\left(\frac{d\tau}{d\theta} + \frac{db}{d\theta}\right)^2}{E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right]}$$

Inoltre, poichè  $\int_{\tilde{\Omega}_x} d\vec{X} L \partial \ln L / \partial \theta = 0$ , derivando rispetto a  $\theta$  si ottiene

$$\int_{\tilde{\Omega}_x} d\vec{X} \left[ \frac{\partial L}{\partial \theta} \frac{\partial \ln L}{\partial \theta} + L \frac{\partial^2 \ln L}{\partial \theta^2} \right] = \int_{\tilde{\Omega}_x} d\vec{X} L \left[ L \left( \frac{\partial \ln L}{\partial \theta} \right)^2 + L \frac{\partial^2 \ln L}{\partial \theta^2} \right] = E \left[ \left( \frac{\partial \ln L}{\partial \theta} \right)^2 \right] - E \left[ - \frac{\partial^2 \ln L}{\partial \theta^2} \right] = 0$$

da cui eq. (21).

La quantità  $I = E \left[ -\partial^2 \ln L / \partial \theta^2 \right]$ , legata alle varianze, è detta quantità di informazione o informazione di Fisher.

Notare che

- il  $\ln L$  deve essere derivabile due volte rispetto a  $\theta$ ;
- se  $b = 0$  eq. (21) ha un'espressione più semplice e se, come spesso accade, è anche  $\tau(\theta) = \theta$ , la disuguaglianza diventa

$$\sigma_t^2 \geq \frac{1}{E \left[ - \frac{\partial^2 \ln L}{\partial \theta^2} \right]} \quad (22)$$

- l'uguaglianza vale se e solo se

$$\frac{\partial \ln L}{\partial \theta} = k(t - \tau) \quad (23)$$

dove  $k$  può anche essere funzione di  $\theta$ .

Se  $t$  è tale per cui vale l'uguaglianza, lo stimatore si dice "a varianza minima". In generale si definisce come "efficienza" di uno stimatore il rapporto  $\epsilon_t = \sigma_{min}^2 / \sigma_t^2$ , che è uguale a 1 solo per stimatori a varianza minima, che utilizzano tutte le informazioni fornite dai campioni.

\* esempi di calcolo della varianza e della varianza minima

- stima del valore di aspettazione della distribuzione di Poisson (vedi appunti);
- stima del valore di aspettazione della distribuzione di Gauss (vedi appunti);
- stima della varianza della distribuzione  $N(0, \sigma^2)$  (vedi appunti);
- stima della deviazione standard della distribuzione  $N(0, \sigma^2)$  (da fare come esercizio).

Tra le altre proprietà di cui gli stimatori dovrebbero godere c'è la "sufficienza": uno stimatore è sufficiente se non possono essere ottenute altre informazioni sul parametro a partire dal campione. Tutti gli stimatori efficienti sono anche sufficienti.

Per quanto riguarda l'espressione esplicita degli stimatori, finora sono stati considerati alcuni esempi semplici. In generale, come già detto, esistono dei metodi che possono essere utilizzati per identificare stime dei parametri con le proprietà viste, che sono descritti nei prossimi due paragrafi.

### 3.3 Metodo del Maximum Likelihood

Questo metodo (MML), molto generale e diffuso, permette di determinare stimatori con proprietà richieste. In condizioni molto generali si può dimostrare che le stime ottenute con il MML sono consistenti. Sono inoltre sempre asintoticamente centrate e spesso centrate. Infine, almeno asintoticamente, sono efficienti e con funzione di distribuzione di Gauss.

#### 3.3.1 Stimatori

Il principio su cui il MML si basa può essere riassunto nel seguente modo. Consideriamo una variabile casuale  $X$  con funzione di distribuzione  $f(X, \theta)$ , dove  $\theta$  è il parametro da stimare a partire da un insieme di misure indipendenti  $X_1, X_2, \dots, X_n$ . Per un certo  $\theta$  la funzione densità di probabilità congiunta per le variabili  $X_i$  è

$$L(\vec{X}, \theta) = \prod_{i=1}^n f(X_i, \theta). \quad (24)$$

che può anche essere vista come la funzione densità di probabilità del parametro  $\theta$  per valori fissati delle variabili casuali  $X_i = x_i$ , cioè quando calcolata su un campione specifico. Il principio del Maximum Likelihood afferma che la stima migliore di  $\theta$  da un certo campione è il valore  $\hat{\theta}$  che massimizza la funzione  $L$ .

Spesso  $\hat{\theta}$  può essere determinato analiticamente. Per avere un massimo deve infatti essere:

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{oppure} \quad \frac{\partial \ln L}{\partial \theta} = 0 \quad (25)$$

equazione che permette di determinare l'espressione di  $t(\vec{X})$  il cui valore di aspettazione è  $E[t] = \theta$  e il cui valore numerico, quando calcolato sul campione  $x_1, x_2, \dots, x_n$ , è la stima del parametro  $\hat{\theta} = t(\vec{x})$ . Naturalmente bisogna verificare che sia

$$\left[ \frac{\partial^2 \ln L}{\partial \theta^2} \right]_{t=E[t]} < 0. \quad (26)$$

L'estensione al caso di  $k$  parametri è ovvia: si tratta di determinare le  $k$  funzioni del campione  $t_j(\vec{X})$  soluzioni delle  $k$  equazioni

$$\frac{\partial \ln L}{\partial \theta_j} = 0. \quad (27)$$

Nel caso in cui non esista un'unica soluzione, sono necessarie ulteriori informazioni (il campione non è sufficiente a risolvere il problema), mentre se la soluzione analitica non può essere ottenuta facilmente si può massimizzare  $L$  o  $\ln L$  numericamente (come fatto dai programmi di fit più diffusi, come Minuit).

### 3.3.2 Esempi

Un esempio interessante è la stima della vita media. In questo caso il campione è costituito dalla misura di  $n$  intervalli di tempo  $x_i$  tra un istante arbitrario e l'istante in cui si verifica l'evento. L'ipotesi è che la funzione di distribuzione degli intervalli di tempo sia

$$f(X, \tau) = \frac{1}{\tau} e^{-X/\tau} \quad (28)$$

dove  $\tau$  è la vita media che si vuole stimare usando il campione a disposizione. La funzione di likelihood è quindi

$$L(\vec{X}, \tau) = \frac{1}{\tau^n} \prod_{i=1}^n e^{-X_i/\tau}. \quad (29)$$

Ponendo la derivata di  $\ln L$  rispetto a  $\tau$  uguale a 0 si ottiene che lo stimatore di  $\tau$  è

$$t_\tau = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (30)$$

ed è

$$E \left[ \frac{\partial^2 L}{\partial \theta^2} \right]_{t=E[t]} = E \left[ \frac{n}{\tau^3} (\tau - 2n\bar{X}) \right]_{t=E[t]} = -\frac{n}{\tau^2} < 0, \quad (31)$$

essendo  $E[X_i] = \tau$ .

Se si fosse ipotizzata la funzione di distribuzione

$$f(X, \lambda) = \lambda e^{-\lambda X} \quad (32)$$

seguendo lo stesso procedimento avremmo ottenuto come stimatore di  $\lambda$  la funzione  $t_\lambda = 1/\bar{X} = 1/t_\tau$ , del tutto ragionevole essendo  $\lambda = 1/\tau$ .

In realtà questa è una proprietà generale degli stimatori ottenuti con il MML. Infatti, se interessa la stima di una generica funzione  $a(\theta)$ , nel caso  $\partial a/\partial \theta \neq 0$ , gli stimatori  $t_\theta$  e  $t_a$  si ottengono rispettivamente da

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \ln L}{\partial a} \frac{\partial a}{\partial \theta} = 0 \quad e \quad \frac{\partial \ln L}{\partial a} = 0,$$

per cui deve essere

$$t_a = a(t_\theta).$$

In generale però  $E[a(t_\theta)] \neq a(E[t_\theta])$  per cui non è detto che entrambe le stime siano centrate.

Tornando agli stimatori di  $\tau$  e  $\lambda$  si ha

$$E[\bar{X}] = \tau$$

e quindi è centrato, mentre usando la funzione generatrice dei momenti si ottiene

$$E\left[\frac{1}{\bar{X}}\right] = \frac{n}{n-1}\lambda$$

e lo stimatore è asintoticamente centrato.

Infine, si può verificare facilmente che  $t_\tau$  è a varianza minima.

Altri esempi interessanti nel caso di funzioni di distribuzione di Gauss, già considerati, sono:

- stima di  $\mu$  nel caso di stesso valore di aspettazione e stessa varianza nota ( $N(\mu, \sigma^2)$ )
- stima di  $\mu$  nel caso di stesso valore di aspettazione ma varianze diverse e note ( $N(\mu, \sigma_i^2)$ )
- stima di  $\sigma^2$  nel caso di stesso valore di aspettazione noto e stessa varianza ( $N(\mu, \sigma^2)$ )
- stima simultanea di  $\mu$  e  $\sigma^2$  nel caso di stesso valore di aspettazione e stessa varianza ( $N(\mu, \sigma^2)$ )

\* vedi appunti

### Esercizi

1. Determinare con il MML la stima di  $\tau$  nel caso in cui gli intervalli di tempo misurati siano limitati all'intervallo  $(0, T)$ .

2. Dimostrare che:

- le stime di  $\mu$  e  $\sigma^2$  corrispondono effettivamente a un massimo di  $L$ ;
- la stima di  $\sigma$  è la radice quadrata della stima di  $\sigma^2$ ;
- la stima di  $\sigma$  non è centrata ma è consistente.

### 3.3.3 Ulteriori considerazioni

#### Proprietà delle stime

Il MML è molto usato per le proprietà di cui godono le stime:

- in condizioni del tutto generale, nel caso di uno o più parametri, le stime sono consistenti.
- le stime sono asintoticamente centrate e spesso centrate, cioè  $E[t] = \theta$ , indipendentemente dalla dimensione del campione. Se non sono centrate, si può applicare una correzione per eliminare il bias (come nel caso della stima della varianza).
- le stime sono almeno asintoticamente efficienti e generalmente per la varianza della stima di un parametro  $\theta$  si usa l'espressione

$$\sigma_{\hat{\theta}}^2 = \frac{1}{\left[-\frac{\partial^2 \ln L}{\partial \theta^2}\right]_{\theta=\hat{\theta}}} \quad (33)$$

- proprietà molto importante delle stime è che almeno asintoticamente hanno funzione di distribuzione di Gauss.

### Stima di più parametri

Nel caso di stime di più parametri, assumendo che siano efficienti e centrati, gli stimatori per i diversi parametri si ottengono risolvendo il sistema di equazioni (27). Gli elementi della matrice  $V$  delle covarianze si ottengono da

$$(V^{-1})_{ij} = \left[ -\frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_k} \right]_{\vec{\theta} = \vec{\hat{\theta}}} \quad (34)$$

### Metodo grafico/numerico

Nel caso in cui le soluzioni analitiche siano difficili da calcolare, come spesso succede, si possono utilizzare metodi numerici. Asintoticamente, infatti, le stime sono gaussiane, e, per il significato della funzione di Likelihood, deve essere:

$$L(t) = L_0 e^{-\frac{(t-\theta_0)^2}{2\sigma_t^2}} \quad (35)$$

dove  $t$  è lo stimatore del parametro e  $\theta_0$  il suo valore. Data la stima  $\hat{\theta}$ , i possibili valori delle stime hanno quindi distribuzione

$$L(\theta) = L_{max} e^{-\frac{(\theta-\hat{\theta})^2}{2\sigma_{\hat{\theta}}^2}} \quad (36)$$

dove  $L_{max}$  è definito dal campione specifico utilizzato per determinare  $\hat{\theta}$ . Da qui si ottiene che

$$\ln L(\theta) = \ln L_{max} - \frac{(\theta - \hat{\theta})^2}{2\sigma_{\hat{\theta}}^2} \quad (37)$$

cioè il  $\ln L$  ha asintoticamente un andamento parabolico in funzione dei possibili valori del parametro con un massimo per  $\theta = \hat{\theta}$ , dove vale  $\ln L_{max}$ . Il valore  $\ln L_{max} - 1/2$  si ottiene per

$$\theta - \hat{\theta} = \pm \sigma_{\hat{\theta}}. \quad (38)$$

Per ottenere la stima del parametro e la sua deviazione standard si può quindi procedere graficamente (o numericamente): si calcola  $\ln L$  sul campione per diversi valori di  $\theta$ , si riportano in un grafico i valori ottenuti e si determina la parabola. Il valore massimo individua la stima  $\hat{\theta}$  e i valori di  $\theta$  corrispondenti a  $\ln L_{max} - 1/2$  permettono di determinare  $\sigma_{\hat{\theta}}$ , cioè l'intervallo corrispondente a un livello di confidenza di 0.68 (distribuzione di Gauss), i valori di  $\theta$  corrispondenti a  $\ln L_{max} - 2$  permettono di determinare  $2\sigma_{\hat{\theta}}$ , ecc.. Notare che la curva ottenuta potrebbe non essere esattamente una parabola nel caso in cui le proprietà delle stime siano solo asintotiche.

L'eq. (37) si può ottenere anche sviluppando in serie  $\ln L$  in un intorno della stima:

$$\ln L(\theta) = \ln L(\hat{\theta}) + \left[ \frac{\partial \ln L}{\partial \theta} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2} \left[ \frac{\partial^2 \ln L}{\partial \theta^2} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \dots \quad (39)$$

Il primo termine è  $\ln L_{max}$ , il secondo è nullo perchè la stima corrisponde al massimo di  $\ln L$  ed il terzo è il secondo di eq. (37), nell'ipotesi si stime efficienti.

Nel caso di  $k$  parametri, generalizzando quanto visto al caso di variabili con funzione di distribuzione multinormale, eq. (37) diventa

$$\ln L(\vec{\theta}) = \ln L_{max} - \frac{1}{2}(\vec{\theta} - \vec{\hat{\theta}})^T V^{-1}(\vec{\theta} - \vec{\hat{\theta}}) \quad (40)$$

Il valore  $\ln L_{max}$  corrisponde alle stime  $\vec{\hat{\theta}}$ . I valori di  $\ln L(\vec{\theta}) = \ln L_{max} - 1/2$  permettono di determinare la regione di confidenza nello spazio a  $k$  dimensioni dei parametri che corrisponde a una deviazione standard per ciascun parametro e che contiene i valori veri dei parametri con probabilità data da  $P(\chi^2_{\nu=k} \leq 1)$ . In questo caso la procedura è più complessa e si affronta in modo iterativo (o costruendo una griglia a  $k$  dimensioni sui cui punti calcolare  $\ln L$ ). Per  $k = 2$ , si fissano diversi valori ad esempio di  $\theta_1$  e si determina per ciascuno la parabola in funzione di  $\theta_2$ . I massimi di  $\ln L$ , riportati in funzione di  $\theta_1$  si trovano ancora su una parabola il cui massimo è  $\ln L_{max}$ . I punti corrispondenti a  $\ln L_{max} - 1/2$  appartengono a un'ellisse, la già introdotta ellisse di confidenza, che permette di determinare varianze e covarianza.

### Binned e Unbinned Maximum Likelihood

Avendo a disposizione un numero elevato di misure di una variabile casuale  $x$  la cui funzione di distribuzione dipende dai parametri da stimare, si può procedere in modo diverso. Finora si è considerata la funzione di Likelihood delle singole misure (Unbinned ML). In alternativa si può costruire l'istogramma delle misure e considerare, come campione, il numero di misure  $y_i$  in ciascuno degli  $m$  intervalli dell'istogramma (Binned ML). La funzione di Likelihood è una funzione multinomiale e i valori di aspettazione dei numeri di eventi nei diversi bin sono dati dalla funzione di distribuzione di  $x$  e quindi dipendono dai parametri. Scritta la funzione  $L$  per il nuovo campione di dimensione  $m$  si procede poi come nel caso dell'UML. Da notare che  $L$  sarà diversa a seconda che si consideri il numero totale di eventi una variabile casuale o un numero fissato.

\* vedi appunti per esempi

## 3.4 Metodo dei Minimi Quadrati

### 3.4.1 Principio

Anche questo metodo (MMQ o "least squared method", LSM) permette di ottenere stime con molte delle proprietà auspicabili elencate in precedenza. In particolare, per problemi lineari gli stimatori sono a varianza minima e senza bias anche nel caso di campioni di dimensione piccola, e possono essere calcolati analiticamente.

Il metodo è descritto in modo sintetico nel seguito, per il caso di una serie di  $N$  coppie di misure indipendenti di due grandezze fisiche  $Z$  e  $Y$ . Supponiamo che tra le grandezze fisiche ci sia una dipendenza funzionale (nota) del tipo  $Y = f(Z, \vec{\theta})$ , dove  $\vec{\theta}$  è il vettore dei  $k$  parametri che si vogliono stimare. Avendo eseguito  $N > k$  misure, per  $N$  valori di  $Z$  ( $z_1, z_2, \dots, z_N$ ) avremo  $N$  valori misurati di  $Y$  ( $y_1, y_2, \dots, y_N$ ) i cui valori di aspettazione

sono  $\mu_i = f(z_i, \vec{\theta})$ . Il Principio dei Minimi Quadrati afferma che i valori più attendibili (le stime migliori) dei parametri in base al campione sono quelli che minimizzano la forma quadratica

$$X^2 = \sum_{i=1}^N w_i (y_i - \mu_i)^2 \quad (41)$$

dove  $w_i$  sono i pesi associati alla  $i$ -esima misura e sono legati alle incertezze su  $y_i$ . Se le varianze sono note in genere si pone  $w_i = 1/\sigma_{y_i}^2$ . I valori dei parametri  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$  per i quali  $X^2$  ha il valore minimo sono gli stimatori dei parametri del MMQ. Il principio del MMQ è facilmente comprensibile ed intuitivo: si determinano le stime dei parametri in modo da minimizzare la somma dei quadrati delle differenze tra valori misurati e valori attesi, in unità di deviazioni standard. Come nel caso del MML, se non è possibile trovare un soluzione analitica, si possono usare metodi grafici e/o numerici.

\* esempio:

supponiamo di sapere che il periodo di oscillazione di un pendolo semplice  $T$  è legato alla lunghezza del pendolo  $l$  dalla relazione  $T = cl^{1/2}$  e di voler determinare la costante di proporzionalità  $c$  avendo a disposizione  $N$  misure di  $l$  e  $T$ . In base al MMQ lo stimatore di  $c$  si ottiene minimizzando

$$X^2 = \sum_{i=1}^N \frac{(T_i - cl_i^{1/2})^2}{\sigma_i^2}$$

dove  $\sigma_i^2$  è la varianza di  $T_i$ . Ponendo la derivata di  $X^2$  rispetto a  $c$  uguale a zero, si ottiene

$$\hat{c} = \frac{\sum_{i=1}^N T_i l_i^{1/2} / \sigma_i^2}{\sum_{i=1}^N l_i / \sigma_i^2}.$$

Notare che si è assunto che le incertezze sulle lunghezze siano trascurabili.

Il MMQ può essere usato anche nel caso in cui si vogliono determinare i parametri di una funzione di distribuzione a partire dall'istogramma dei valori assunti da una variabile casuale  $Z$  in  $n$  prove. In questo caso, considerando il generico  $i$ -esimo intervallo dell'istogramma,  $z_i$  è il valore medio dell'intervallo,  $y_i$  è il numero di eventi in quell'intervallo,  $\mu_i$  il suo valore di aspettazione da calcolare a partire dalla funzione di distribuzione di  $Z$  e la varianza di  $y_i$  si ottiene, al solito, con la distribuzione binomiale o di Poisson. C'è da notare, però, che così facendo i parametri compaiono anche al denominatore di  $X^2$ ; per semplificare il problema spesso si assume  $\sigma_i^2 = y_i$ .

Se le varianze  $\sigma_i^2$  sono tutte uguali tra loro, la forma quadratica da minimizzare è semplicemente  $X^2 = \sum_{i=1}^N (y_i - \mu_i)^2$ .

Se le misure  $y_i$  non sono indipendenti, bisogna considerare anche le covarianze, e la forma quadratica da utilizzare è

$$X^2 = (\vec{y} - \vec{\mu})^T V^{-1} (\vec{y} - \vec{\mu}) \quad (42)$$

dove  $V$  è la matrice delle covarianze delle  $\vec{y}$ .

Altri punti importanti sono i seguenti:

- si è assunto che le incertezze sui valori  $z_i$  siano trascurabili. Questo equivale a chiedere che l'incertezza su  $\mu_i$  dovuta all'incertezza su  $z_i$  sia molto minore dell'incertezza

su  $y_i$ . Se questo non è vero, ma sono trascurabili le incertezze su  $y_i$ , si possono semplicemente scambiare i ruoli di  $\vec{z}$  e  $\vec{y}$ . Infine, se le incertezze su entrambe le variabili non sono trascurabili, mantenendo il ruolo iniziale di  $\vec{z}$  e  $\vec{y}$ , si possono sostituire le  $\sigma_{y_i}^2$  con  $\sigma_i^2 = \sigma_{y_i}^2 + \sigma_{\mu_i}^2$  dove  $\sigma_{\mu_i}^2$  è la varianza di  $\mu_i$  ottenuta da  $\sigma_{z_i}^2$  con la legge di propagazione della varianza. Le varianze  $\sigma_{\mu_i}^2$  dipendono però dai valori dei parametri, ed è necessario iterare la procedura di stima dei parametri.

- con il MMQ non si fa alcuna ipotesi sulla funzione di distribuzione delle  $y_i$  ed il metodo è valido per qualsiasi funzione di distribuzione. Nel caso importante in cui le  $y_i$  hanno funzione di distribuzione gaussiana, se l'ipotesi in base alla quale si calcola  $\vec{\mu}$  è corretta, la forma quadratica  $X_{min}^2$  è una variabile casuale con funzione di distribuzione di  $\chi^2$  ed è per questo motivo che il metodo viene anche chiamato minimizzazione del  $\chi^2$ .
- se le  $y_i$  hanno funzione di distribuzione gaussiana, gli stimatori dal MMQ sono gli stessi di quelli che si ottengono con il MML.

### 3.4.2 Il metodo dei Minimi Quadrati Lineare

Un caso particolarmente importante è quello in cui  $\mu_i$ ,  $i = 1, n$ , i valori di aspettazione delle misure  $y_i$ , dipendono linearmente dai  $k$  parametri che si vogliono stimare:

$$\mu_i = f(z_i, \vec{\theta}) = \sum_{j=1}^k a_{ij}(z_i) \theta_j, \quad \text{o } \vec{\mu} = A\vec{\theta}, \quad (43)$$

e i pesi (le varianze  $\sigma_{y_i}^2$ ) non dipendono dai parametri. Il MMQ in questo caso specifico è detto MMQ Lineare (MMQL). In questo caso le stime hanno proprietà importanti: sono uniche, centrate, a varianza minima e asintoticamente gaussiane (gaussiane se le  $y_i$  hanno funzione di distribuzione gaussiana), e possono essere calcolate analiticamente.

Nel seguito è descritto il MMQL nel caso di due parametri e nel caso generale.

#### Parametri di una retta

Particolarmente semplice è il caso in cui la relazione tra le grandezze fisiche è lineare:  $Y = mZ + q$ , con due soli parametri ( $m$  e  $q$ ) da stimare a partire da  $n$  coppie di misure indipendenti  $z_i, y_i$  con incertezze trascurabili per  $z_i$  e incertezza pari a  $\sigma_i^2$  per  $y_i$ . In questo caso la funzione dei parametri da minimizzare è

$$X^2 = \sum_{i=1}^n \frac{[y_i - (mz_i + q)]^2}{\sigma_i^2}. \quad (44)$$

Chiedendo che le derivate rispetto a  $m$  e a  $q$  siano uguali a zero si ottiene facilmente

$$\hat{m} = \frac{S_{00}S_{11} - S_{10}S_{01}}{S_{00}S_{20} - S_{10}^2}, \quad \hat{q} = \frac{S_{20}S_{01} - S_{10}S_{11}}{S_{00}S_{20} - S_{10}^2} \quad (45)$$

dove

$$S_{00} = \sum_{i=1}^n \frac{1}{\sigma_i^2}, \quad S_{01} = \sum_{i=1}^n \frac{y_i}{\sigma_i^2}, \quad S_{10} = \sum_{i=1}^n \frac{z_i}{\sigma_i^2}, \quad S_{11} = \sum_{i=1}^n \frac{z_i y_i}{\sigma_i^2}, \quad S_{20} = \sum_{i=1}^n \frac{z_i^2}{\sigma_i^2}. \quad (46)$$

Infine, con la legge di propagazione della varianza, derivando gli stimatori rispetto a  $y_j$ , si ottiene

$$\sigma_{\hat{m}}^2 = \frac{S_{00}}{S_{00}S_{20} - S_{10}^2}, \quad \sigma_{\hat{q}}^2 = \frac{S_{20}}{S_{00}S_{20} - S_{10}^2}, \quad \text{cov}(\hat{m}, \hat{q}) = -\frac{S_{10}}{S_{00}S_{20} - S_{10}^2}. \quad (47)$$

### Caso generale

Per risolvere nel caso generale un problema con il MMQL è ovviamente necessario identificare bene le quantità rilevanti, e in particolare la matrice  $A$  che permette di ottenere i valori di aspettazione  $\vec{\mu}$  delle  $n$  grandezze fisiche  $\vec{y}$  con matrice delle covarianze  $V$  a partire dai parametri  $\vec{\theta}$  e dalle grandezze fisiche  $\vec{z}$ .

\* esempio:

se la dipendenza di  $Y$  da  $Z$  è del tipo  $Y = a_0 + a_1Z + a_2Z^2$  si ha

$$A = \begin{pmatrix} 1 & z_1 & z_1^2 \\ 1 & z_2 & z_2^2 \\ \vdots & \vdots & \vdots \\ 1 & z_n & z_n^2 \end{pmatrix} \quad \vec{\theta} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix},$$

così che  $\vec{\mu} = A\vec{\theta}$ .

La forma quadratica da minimizzare è data in eq. (42) e può essere riscritta come

$$X^2 = (\vec{y} - A\vec{\theta})^T V^{-1} (\vec{y} - A\vec{\theta}). \quad (48)$$

Derivando  $X^2$  rispetto a  $\vec{\theta}$  e ponendo la derivata uguale a zero si ottiene

$$\vec{\hat{\theta}} = B\vec{y} \quad \text{con} \quad B = (A^T V^{-1} A)^{-1} A^T V^{-1} \quad (49)$$

\* infatti:

$$\frac{\partial X^2}{\partial \vec{\theta}} = -2(\vec{y} - A\vec{\theta})^T V^{-1} A. \quad (50)$$

Da  $[(\vec{y} - A\vec{\theta})^T V^{-1} A]^T = 0$  e ricordando che  $V$  è una matrice simmetrica si ottiene  $A^T V^{-1} \vec{y} = A^T V^{-1} A \vec{\theta}$  e quindi eq.(49).

La matrice  $B$  può essere calcolata se si è in grado di calcolare le matrici inverse. Per quanto riguarda le incertezze e correlazioni tra i parametri, è  $V_{\vec{\theta}} = (A^T V^{-1} A)^{-1}$  e  $B$  può anche essere scritta nella forma

$$B = V_{\vec{\theta}} A^T V^{-1}. \quad (51)$$

\* dimostrazione

### Esercizi

- utilizzare il metodo generale per ottenere le stime dei parametri dei casi  $Y = mZ + q$  e  $Y = mZ$ .

- assumendo che la relazione tra  $Z$  e  $Y$  sia  $Y = a_0 + a_1Z + a_2Z^2$ , stimare i parametri  $a_i$  avendo a disposizione le coppie di valori  $z_i, y_i$ :  $(-0.6, 5.)$ ,  $(-0.2, 3.)$ ,  $(+0.2, 5.)$ ,  $(+0.6, 8.)$ , con deviazioni standard dei valori  $y_i$  pari a 2., 1., 1., 2. e covarianze nulle. Riportare in un grafico valori misurati e curva ottenuta. Commentare i valori numerici dei parametri e le corrispondenti incertezze. Eseguire il test d'ipotesi.
- determinare lo stimatore di una grandezza fisica  $G$  misurata da due diversi esperimenti che hanno ottenuto i valori  $g_A \pm \sigma_A$  e  $g_B \pm \sigma_B$ .
- un esperimento ( $A$ ) ha misurato due grandezze  $G_1$  e  $G_2$  ottenendo  $g_{1A} = 0.12$  nelle opportune unità e  $g_{2A} = -0.25$  con matrice delle covarianze

$$V_A = \begin{pmatrix} 0.01 & -0.01 \\ -0.01 & 0.04 \end{pmatrix}.$$

Un secondo esperimento ( $B$ ) ha misurato solo  $G_1$  ottenendo  $g_{1B} = 0.01 \pm 0.08$ . Calcolare le stime di  $G_1$  e  $G_2$  e corrispondenti varianze e covarianza. Commentare il risultato per  $G_2$ .

### 3.4.3 Incertezze statistiche

Nel caso lineare, facendo riferimento alla notazione dei paragrafi precedenti, abbiamo visto che le stime si ottengono minimizzando la forma quadratica di eq. (48) e possono essere scritte come

$$\vec{\theta} = V_{\vec{\theta}} A^T V^{-1} \vec{y}. \quad (52)$$

Scrivendo  $(\vec{y} - A\vec{\theta}) = (\vec{y} - A\vec{\hat{\theta}}) + A(\vec{\hat{\theta}} - \vec{\theta})$  la forma quadratica diventa

$$\begin{aligned} X^2 &= (\vec{y} - A\vec{\hat{\theta}})^T V^{-1} (\vec{y} - A\vec{\hat{\theta}}) + (\vec{\hat{\theta}} - \vec{\theta})^T A^T V^{-1} A (\vec{\hat{\theta}} - \vec{\theta}) + \\ &+ (\vec{y} - A\vec{\hat{\theta}})^T V^{-1} A (\vec{\hat{\theta}} - \vec{\theta}) + (\vec{\hat{\theta}} - \vec{\theta})^T A^T V^{-1} (\vec{y} - A\vec{\hat{\theta}}). \end{aligned} \quad (53)$$

Il primo termine è il valore minimo di  $X^2$

$$X_{min}^2 = (\vec{y} - A\vec{\hat{\theta}})^T V^{-1} (\vec{y} - A\vec{\hat{\theta}}), \quad (54)$$

il secondo è  $(\vec{\hat{\theta}} - \vec{\theta})^T V_{\vec{\theta}}^{-1} (\vec{\hat{\theta}} - \vec{\theta})$ , mentre il terzo e il quarto contengono la derivata di  $X^2$  rispetto a  $\vec{\theta}$  calcolata per  $\vec{\theta} = \vec{\hat{\theta}}$  che deve essere uguale a zero e quindi sono uguali a zero. Si ha quindi

$$X^2 = X_{min}^2 + (\vec{\hat{\theta}} - \vec{\theta})^T V_{\vec{\theta}}^{-1} (\vec{\hat{\theta}} - \vec{\theta}), \quad (55)$$

espressione che permette di determinare immediatamente le regioni di confidenza. Così per  $X^2 = X_{min}^2 + 1$  si ottiene l'ellissoide con semiassi pari a una deviazione standard, che avrà una probabilità di contenere il valore del parametro data dalla distribuzione

di  $\chi_\nu^2$  con  $\nu = n - k$  se  $n$  sono le misure utilizzate e  $k$  i parametri stimati. Lo stesso risultato si può ottenere sviluppando  $X^2$  in serie in un intorno di  $\vec{\hat{\theta}}$ . Nel caso dei MMQL l'espressione è esatta, essendo le derivate seconde rispetto ai parametri uguali a zero. Nel caso non lineare l'eq. (55) è un paraboloide in genere deformato, così come gli ellipsoidi di confidenza. Nonostante ciò l'approssimazione è generalmente buona in un intorno delle stime e questo permette di determinare stime e regioni di confidenza anche nel caso non lineare. Si possono usare, infatti, metodi numerici e/o grafici analoghi a quelli descritti nel caso del MML per determinare sia le stime, corrispondenti a  $X_{min}^2$ , sia le regioni di confidenza, date da  $X^2 = X_{min}^2 + 1, +4, \dots$

## 4 Intervalli di confidenza

Se lo scopo di un esperimento è determinare un parametro  $\theta$  incognito, quanto già visto è sufficiente per ottenere nei casi più semplici il valore numerico. Questo è il caso misure con distribuzione di Gauss: da quanto fatto, è chiaro come un campione di misure possa essere usato per stimare valore di aspettazione e varianza.

La stima puntuale non è sufficiente però sufficiente, ed è necessario quantificare anche l'errore (nel senso di incertezza) sul valore numerico dovuto alla precisione statistica dell'esperimento. Ciò viene fatto specificando l'intervallo di confidenza che, in molti casi, può essere ottenuto in modo semplice dalla stima della deviazione standard della stima del parametro. Se però la stima non ha funzione di distribuzione di Gauss questo non è il caso ed è necessario usare una procedura più generale. Questo capitolo è dedicato all'introduzione della procedura generale e alla sua applicazione al caso, molto comune in settori diversi, di misure con distribuzione gaussiana. Il caso di distribuzione multinormale verrà trattato nel capitolo sulla stima dei parametri.

Gli "intervalli di confidenza" (o, per più parametri, le "regioni di confidenza") sono costruiti in modo tale da includere il valore vero del parametro con una probabilità maggiore (variabili discrete) o uguale (variabili continue) a un certo valore. Va detto subito che l'intervallo di confidenza non è fisso ma dipende dal campione utilizzato: ripetendo l'esperimento e quindi usando un campione diverso, l'intervallo cambierebbe ogni volta.

Un metodo generale per determinare l'intervallo di confidenza per un parametro è il seguente. Supponiamo che a partire da un campione di dimensione  $n$  si voglia stimare il parametro  $\theta$  e che venga utilizzata come stimatore la statistica  $t$ , funzione del campione, il cui valore di aspettazione sia il valore vero  $\theta_0$  del parametro (o una sua funzione). La funzione di distribuzione di  $t$ ,  $f(t)$ , che supponiamo nota, dipende dal valore del parametro. Avendo fissato un certo valore di probabilità  $\gamma$ , per ogni valore di  $\theta$  si può determinare un intervallo  $[t_a, t_b]$  tale che

$$\int_{t_a}^{t_b} f(t)dt = \gamma \quad (56)$$

Se, dall'esperimento eseguito, si ottengono, fissato  $\gamma$ , i valori  $t_a^*$  e  $t_b^*$ , essendo

$$\gamma = P(t_a^* < t < t_b^*) = P(\theta_a^* < \theta < \theta_b^*) \quad (57)$$

il corrispondente intervallo per  $\theta$ ,  $[\theta_a^*, \theta_b^*]$ , è detto intervallo di confidenza con "livello di confidenza"  $\gamma$ . Il significato è che, se l'esperimento fosse ripetuto un numero elevato di volte, l'intervallo  $[\theta_a^*, \theta_b^*]$  varierebbe includendo il valore del parametro  $\gamma$  volte.

Per quanto riguarda il livello di confidenza, la sua scelta è piuttosto arbitraria. Un valore basso implica un intervallo minore ma con piccolo contenuto di probabilità. In genere si scelgono valori vicini a 1, 0.90 o 0.95 o 0.99, tranne nel caso di funzione di distribuzione normale, in cui spesso si sceglie 0.68.

Notare che la condizione di eq. (56) non determina in modo univoco gli estremi dell'intervallo e per farlo bisogna usare criteri aggiuntivi. Il più comune consiste in scegliere

Intervalli centrali, tali cioè che la probabilità che  $t$  sia minore di  $t_a$  sia uguale alla probabilità che  $t$  sia maggiore di  $t_b$  e uguale a  $(1 - \gamma)/2$ .

Di particolare rilevanza è la costruzione degli intervalli di confidenza si hanno nel caso in cui si abbia a che fare con distribuzioni normali (trattato nel prossimo paragrafo) e multinormali (come nel caso di stime di parametri a partire da campioni sufficientemente grandi: l'argomento verrà ripreso nel capitolo 3).

Il collegamento con il test d'ipotesi (capitolo 5) sarà chiaro nel seguito.

#### 4.1 Misure con distribuzione di Gauss

L'identificazione degli intervalli di confidenza nel caso di un'unica misura con funzione di distribuzione di Gauss è un problema piuttosto comune (misure con errori accidentali ma anche stima di un parametro a partire da un insieme di dati sufficientemente ampio).

Nel seguito è considerato il caso in cui il valore misurato è una variabile casuale con funzione di distribuzione  $N(\mu, \sigma^2)$  dove il valore di aspettazione  $\mu$  sia il valore della grandezza fisica che si vuole determinare e la varianza  $\sigma^2$  sia dovuta alla precisione del metodo di misura. Avendo a disposizione un campione costituito da  $n$  misure  $x_i$  indipendenti, lo stimatore  $\bar{x} = (\sum_i x_i)/n$  ha distribuzione di Gauss  $N(\mu, \sigma^2/n)$ . La stima puntuale è il valore numerico di  $\bar{x}$  sul campione. Per determinare gli intervalli di confidenza, però, è necessario distinguere i casi in cui la varianza è nota oppure stimata dal campione stesso.

##### Varianza nota

La variabile

$$y = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (58)$$

ha distribuzione normale standard e la probabilità che il valore misurato di  $y$  cada in  $(-1, 1)$  (cioè che  $\bar{x}$  differisca da  $\mu$  per meno di  $\pm\sigma$  è il 68.3%. L'intervallo di confidenza  $(\bar{x} - \sigma, \bar{x} + \sigma)$  corrisponde quindi a un livello di confidenza  $\gamma = 0.683$ . Analogamente l'intervallo  $(\bar{x} - 2\sigma, \bar{x} + 2\sigma)$  corrisponde a  $\gamma = 0.9545$  e così via. Viceversa, si può fissare ad esempio  $\gamma = 0.90$  e determinare, dai valori tabulati della distribuzione normale standard, l'intervallo di confidenza corrispondente.

Gli stessi valori si ottengono considerando  $y^2$ , che è una variabile casuale con distribuzione di  $\chi_1^2$ .

##### Varianza non nota

Se, oltre a  $\mu$ , anche il valore di  $\sigma^2$  è stimato a partire dal campione, la costruzione degli intervalli di confidenza per  $\mu$  è più complicata. Avendo usato, per stimare  $\sigma^2$ :

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 \quad (59)$$

possiamo introdurre, oltre a  $y$  definita in eq.(58), la variabile causale

$$z = \frac{(n-1)s^2}{\sigma^2} \quad (60)$$

che è una variabile casuale di  $\chi_{n-1}^2$ . Il rapporto tra una variabile casuale con distribuzione  $N(0, 1)$  e la radice quadrata di una variabile casuale con distribuzione  $\chi_{\nu}^2$  divisa per il numero di gradi di libertà, è una variabile casuale ( $t$ ) con funzione di distribuzione nota. Si tratta della funzione di distribuzione “t di Student” a  $\nu$  gradi di libertà:

$$f(t, \nu) = \frac{\Gamma[(\nu + 1)/2]}{\sqrt{\pi\nu}\Gamma[\nu/2]} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}. \quad (61)$$

Il valore di aspettazione è  $E[t] = 0$  e, per  $\nu > 2$ , la varianza è  $\sigma_t^2 = \nu/(\nu - 2)$ . Per  $\nu = 1$  la funzione di distribuzione è la funzione di distribuzione di Cauchy, mentre per  $\nu \rightarrow \infty$  è la funzione di distribuzione di Gauss.

Con riferimento alle eq. (58) e (60), possiamo quindi introdurre la variabile casuale

$$t = \frac{y}{\sqrt{z/(n-1)}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad (62)$$

che ha distribuzione t di Student a  $n - 1$  gradi di libertà di cui la funzione cumulativa è tabulata. Si segue quindi la stessa procedura del caso precedente per determinare gli intervalli di confidenza.

Da notare che nella definizione di  $t$  non compare più esplicitamente  $\sigma$ , che quindi non deve essere determinato per ottenere gli intervalli di confidenza per  $\mu$ . Inoltre, per  $n$  sufficientemente grande,  $t$  tende ad avere distribuzione normale standard, e gli intervalli di confidenza conciono con quelli che si otterrebbero seguendo la procedura con varianza nota.

### **Intervalli di confidenza per varianza e deviazione standard**

Anche questo è un problema piuttosto comune in molti campi. Per il procedimento da seguire nel caso di valore di aspettazione noto e non, vedi appunti.

## 5 Test d'ipotesi

Assieme alla stima dei parametri, questo è uno degli argomenti più interessanti dell'inferenza statistica ed è strettamente legato a definizione e significato degli intervalli di confidenza. Anche per questo capitolo verranno dati i concetti di base nell'ambito della statistica classica ed alcuni esempi di metodi e applicazioni.

Dato un fenomeno che coinvolge una o più variabili casuali e un modello statistico ("ipotesi nulla",  $H_0$ ) per la loro funzione di distribuzione (o per la relazione tra le variabili) il "test d'ipotesi" permette di valutare quantitativamente l'accordo tra modello e le osservazioni. Se il test riguarda valori dei parametri è detto parametrico, mentre, se è un test sulla forma della distribuzione (o sulla relazione) per determinati valori dei parametri, è detto non parametrico.

Il test d'ipotesi è una procedura che permette, in base al campione a disposizione, di rigettare o meno l'ipotesi nulla e di valutare la probabilità di errore nel caso in cui l'ipotesi nulla venga rigettata. Tra i diversi test d'ipotesi, bisognerebbe scegliere quello che massimizza la probabilità di rigettare l'ipotesi nulla  $H_0$  se è vera un'ipotesi alternativa  $H_1$ , e minimizza la probabilità di rigettare l'ipotesi nulla se corretta.

Dati l'ipotesi nulla  $H_0$ , l'ipotesi alternativa  $H_1$  e un campione di dimensione  $n$  ( $x_1, \dots, x_n$ ), la procedura generale per il test d'ipotesi può essere riassunta nel seguente modo:

- Si introduce una opportuna statistica, cioè una funzione del campione,  $t(x_1, \dots, x_n)$  che è detta "statistica di test" e della quale nell'ipotesi  $H_0$  (e  $H_1$ ) si conosce la funzione di distribuzione  $\Phi_0(t)$ . Ad esempio, la funzione di distribuzione di  $t$  nel caso in cui sia vera l'ipotesi nulla  $H_0$  potrebbe essere la funzione  $\Phi_0(t)$  riportata in fig. 7. Se invece fosse vera un'ipotesi alternativa  $H_1$  potrebbe essere la funzione  $\Phi_1(t)$ .
- Si sceglie un "livello di significatività"  $\alpha$  per il test d'ipotesi, in genere 0.10, o 0.05, o 0.01. Essendo  $\Phi_0(t)$  nota, dal valore di  $\alpha$  si calcola  $t_\alpha$  tale che

$$\int_{t_\alpha}^{t_{max}} \Phi_0(t) dt = \alpha. \quad (63)$$

cioè tale che la probabilità che  $t$  sia maggiore di  $t_\alpha$  sia  $\alpha$ . L'intervallo  $(t_\alpha, t_{max})$  è detto "regione critica".

- Se, sul campione a disposizione, la statistica di test assume un valore  $t^*$  maggiore di  $t_\alpha$ , cioè se  $t^*$  cade nella regione critica, si rigetta l'ipotesi nulla  $H_0$  con un livello di significatività  $\alpha$ .  $\alpha$  è quindi la probabilità di rigettare l'ipotesi nulla anche se corretta.

Minore è  $\alpha$ , minore è la probabilità di rigettare  $H_0$  se corretta (errore del I tipo). Al diminuire di  $\alpha$  aumenta però la probabilità  $\beta$

$$\beta = \int_{t_{min}}^{t_\alpha} \Phi_1(t) dt,$$

di non rigettare l'ipotesi nulla se è vera l'ipotesi alternativa  $H_1$  (errore del II tipo). Le probabilità di errore nella decisione presa dipendono dal livello di significatività scelto ma



f240114/pdfs-eps-converted-to.pdf

Figura 7: Funzioni di distribuzione  $\Phi(t)$  vs  $t$  per due diverse ipotesi.

anche dalla statistica di test  $t$  utilizzata. In generale per la scelta della statistica di test si dovrebbe seguire il seguente principio: tra tutte le statistiche di test si sceglie quella che, a parità di  $\alpha$ , minimizza  $\beta$ . In pratica non è facile individuare l'ipotesi alternativa  $H_1$  tra le molte possibili e spesso  $H_1$  è identificata con “ $H_0$  falsa” e  $\beta$  non viene calcolato.

Nella descrizione del test d'ipotesi:

- è stato usato il livello di significatività “a priori”. In alternativa si può utilizzare per il livello di significatività “a posteriori” il valore  $\alpha^*$  con  $\alpha^* = \int_{t^*}^{t_{max}} \Phi_o(t) dt$ ;
- è stato considerato il test “a una coda”, avendo introdotto la regione critica  $(t_\alpha, t_{max})$ . Il test a “due code” è trattato negli esempi descritti nel seguito.

## 5.1 Test parametrici per variabili gaussiane

Si tratta di un insieme di test d'ipotesi molto diffusi che coinvolgono i valori dei parametri di una funzione di distribuzione di Gauss. Dato un campione  $x_1, \dots, x_n$  di dimensione  $n$  che si assume rappresentativo di una popolazione con funzione di distribuzione normale  $N(\mu, \sigma^2)$  si vogliono eseguire test sull'accordo tra valori specifici di  $\mu$  e  $\sigma^2$  e le  $n$  osservazioni disponibili. Gli esempi che seguono sono per certi aspetti simili a quanto visto nel paragrafo 4.1.

**Test della media** Nel caso in cui la varianza  $\sigma^2$  sia nota, si vuole verificare se il campione è in accordo con l'ipotesi nulla  $H_o: \mu = \mu_0$  con  $\mu_0$  valore previsto del valore di aspettazione. Come ipotesi alternativa si può scegliere  $H_1: \mu \neq \mu_0$ .

Se  $H_o$  è vera, la media del campione  $\bar{x}$  ha funzione di distribuzione  $N(\mu, \sigma^2/n)$ . Come

statistica di test è conveniente usare

$$t(x_1, \dots, x_n) = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}. \quad (64)$$

che, nell'ipotesi  $H_o$ , ha funzione di distribuzione  $N(0, 1)$ .

Poichè l'ipotesi alternativa è complementare ( $H_1: \mu \neq \mu_0$ ),  $H_o$  va rigettata se  $t$  assume valori grandi, sia positivi che negativi. Fissato il valore di  $\alpha$ , in questo caso è corretto eseguire il test "a due code". La regione critica è costituita da due intervalli simmetrici rispetto a 0, entrambi corrispondenti a un livello di significatività pari a  $\alpha/2$ . Si definisce quindi  $t_{\alpha/2} > 0$  tale che

$$\int_{-\infty}^{-t_{\alpha/2}} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} e^{-t^2/2} dt = \int_{t_{\alpha/2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} e^{-t^2/2} dt = \frac{\alpha}{2} \quad (65)$$

e si rigetta  $H_o$  con livello di significatività  $\alpha$  se il valore di  $t$  per il campione a disposizione è  $t^* < -t_{\alpha/2}$  o  $t^* > t_{\alpha/2}$ .

Per ogni valore scelto di  $\alpha$ ,  $t_{\alpha/2}$  si ottiene usando i valori tabulati della funzione cumulativa di probabilità della funzione di distribuzione normale standard. Ad esempio, se  $\alpha = 0.10$ , è  $t_{\alpha/2} = 1.6$ .

Nel caso in cui la varianza non sia nota, ma venga stimata dal campione, la statistica di test utilizzata ha distribuzione  $t$  di Student.

\* vedi appunti e 4.1

### Test della varianza

Supponiamo che sul valore noto di  $\mu$  non ci siano dubbi e che si voglia verificare l'accordo del campione con il valore previsto della varianza  $\sigma_0^2$ . In questo caso l'ipotesi nulla è  $H_o: \sigma^2 = \sigma_0^2$ . L'ipotesi alternativa può essere  $H_1: \sigma^2 \neq \sigma_0^2$ .

Sapendo che, se  $H_o$  è vera, la statistica

$$t(x_1, \dots, x_n) = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_0^2} \quad (66)$$

ha funzione di distribuzione di  $\chi^2$  con  $n$  gradi di libertà, si può usare  $t$  come statistica di test.

Facendo nuovamente un test a due code, fissato il livello di significatività  $\alpha$ , dai valori tabulati della funzione cumulativa di  $\chi^2$ , si trovano i valori  $t_{\alpha/2}^-$  e  $t_{\alpha/2}^+$  tali che gli integrali della funzione di distribuzione di  $\chi^2$  con  $n$  gradi di libertà fino a  $t_{\alpha/2}^-$  e da  $t_{\alpha/2}^+$  siano entrambi uguali a  $\alpha/2$ . Notare che in genere sarà  $n - t_{\alpha/2}^- \neq t_{\alpha/2}^+ - n$  perchè la funzione di distribuzione di  $\chi^2$  è molto asimmetrica per  $n$  piccolo.

L'ipotesi  $H_o$  verrà rigettata se  $t^*$ , il valore della statistica  $t$  calcolato sul campione a disposizione, cade nella regione critica costituita dagli intervalli  $(0, t_{\alpha/2}^-)$  e  $(t_{\alpha/2}^+, +\infty)$ .

Se l'ipotesi alternativa fosse stata  $H_1: \sigma^2 > \sigma_0^2$ , sarebbe stato opportuno eseguire un test d'ipotesi a una coda con regione critica  $(t_{\alpha}, +\infty)$  in quanto valori piccoli di  $t^*$  indicano

eventualmente che è  $\sigma^2 < \sigma_0^2$ .

Test sulla media e sulla varianza di un campione sono molto diffusi per cui in molti testi di statistica si trova la casistica completa dei possibili test parametrici per variabili gaussiane.

## 5.2 Test di $\chi^2$

È, per la sua semplicità, un test d'ipotesi molto utilizzato. Supponiamo che il campione sia costituito da  $n$  misure indipendenti  $y_i$  con funzione di distribuzione di Gauss con varianze  $\sigma_i^2$  note. Supponiamo di voler verificare che i dati siano in accordo con certi valori di aspettazione  $\mu_i$  di  $y_i$ , ottenuti dalla teoria o da esperimenti precedenti, cioè di voler eseguire un test d'ipotesi con l'ipotesi nulla  $H_o: E[y_i] = \mu_i, i = 1, n$ . Se questo è il caso, la statistica

$$t(y_1, \dots, y_n) = \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2} \quad (67)$$

ha funzione di distribuzione di  $\chi^2$  con  $n$  gradi di libertà e può essere utilizzata come statistica di test. Il disaccordo tra campione e ipotesi statistica è in questo caso caratterizzato da valori 'alti' di  $t$ , mentre valori molto minori di  $n$  possono essere dovuti a una sovrastima delle incertezze, ma non indicano disaccordo con  $H_o$ . Si esegue quindi un test a una coda, definendo in corrispondenza al livello di significatività  $\alpha$  il valore critico  $t_\alpha$  tale che

$$\int_{t_\alpha}^{\infty} f(t) dt = \alpha \quad (68)$$

dove  $f(t)$  è la funzione di distribuzione di  $\chi^2$  con  $n$  gradi di libertà, e l'ipotesi  $H_o$  viene rigettata se il valore  $t^*$  di  $t$  calcolato sul campione cade nella regione critica, cioè se è  $t^* > t_\alpha$ .

Alcune note:

1. Se le variabili  $y_i$  non hanno funzione di distribuzione di Gauss, la statistica  $t$  definita in eq. (67) non ha funzione di distribuzione di  $\chi^2$ . Per usarla come statistica di test bisogna determinarne la funzione di distribuzione, ad esempio simulando molte volte l'esperimento.
2. Se le variabili  $y_i$  hanno funzione di distribuzione di Gauss ma non sono indipendenti, la statistica da usare è

$$t(y_1, \dots, y_n) = (\vec{y} - \vec{\mu})^T V^{-1} (\vec{y} - \vec{\mu}) \quad (69)$$

dove  $V$  è la matrice delle covarianze di  $(\vec{y})$ . Anche in questo caso  $t$  ha funzione di distribuzione di  $\chi^2$  con  $\nu = n$  gradi di libertà.

3. Se i valori  $\mu_i$  sono stati ottenuti utilizzando lo stesso campione usato per il test d'ipotesi,  $t$  ha funzione di distribuzione di  $\chi^2$  con  $\nu < n$  gradi di libertà. È questo il caso in cui i valori  $y_1$  vengono utilizzati per stimare  $k$  parametri incogniti e le stime

vengono poi usate per calcolare i valori  $\mu_i$ : il numero di gradi di libertà è  $\nu = n - k$ . A questo proposito, il valore minimo della forma quadratica  $X^2$  utilizzata per stimare i parametri con il metodo dei minimi quadrati ha la stessa espressione della statistica  $t$  e, nelle stesse condizioni, è una variabile casuale con funzione di distribuzione di  $\chi^2$  e il valore di  $X_{min}^2$  ottenuto permette di eseguire direttamente anche il test d'ipotesi.

4. Se  $\nu$  è sufficientemente grande la funzione di distribuzione di  $\chi^2$  tende a funzione di distribuzione di Gauss con valore di aspettazione  $\nu$  e varianza  $2\nu$ . Questo permette di avere immediatamente un'idea dell'accordo tra campione e modello, senza calcolare  $t_\alpha$ . Inoltre, in media ci si aspetta che ciascun addendo contribuisca al  $\chi^2$  con circa 1: se un contributo è maggiore di 9 (differenza maggiore di  $3\sigma_i$ ) per uno specifico punto il valore di  $y$  corrispondente andrebbe verificato.
5. Nel caso in cui si abbiano a disposizione  $m$  campioni diversi, si può usare una procedura diversa: per ciascun campione si calcola il valore di  $t_j^*$ ,  $j = 1, m$  e si confronta la distribuzione delle probabilità

$$p_j^* = \int_0^{t_j^*} f_j(t) dt$$

dove  $f_j(t)$  è la funzione di distribuzione di  $\chi^2$  con  $\nu_j$  gradi di libertà, con la distribuzione uniforme.

6. Il test di  $\chi^2$  è significativo solo se permette di rigettare l'ipotesi nulla con un certo livello di significatività. Scegliere tra due o più ipotesi alternative, nessuna delle quali possa essere rigettata, sulla base del valore  $t^*$  ottenuto non è corretto.

Il test di  $\chi^2$  ha delle limitazioni tra le quali il fatto di essere poco selettivo se  $\nu$  è piccolo e di non essere sensibile al segno delle differenze tra valore misurato e atteso. D'altra parte può essere applicato con successo in molti casi, come evidente dagli esempi descritti nel seguito.

Supponiamo che  $y_i$  siano misure indipendenti della stessa grandezza fisica e di volerne verificare la compatibilità. In questo caso è  $H_o: E[y_i] = \mu, i = 1, n$ . Se  $\mu$  è noto, la statistica

$$t(y_1, \dots, y_n) = \sum_i \frac{(y_i - \mu)^2}{\sigma_i^2}$$

ha funzione di distribuzione di  $\chi^2$  con  $\nu = n$  gradi di libertà. Se  $\mu$  non è noto, la statistica

$$t(y_1, \dots, y_n) = \sum_i \frac{(y_i - \bar{y})^2}{\sigma_i^2}$$

con  $\bar{y}$  media pesata (o aritmetica, se i valori  $\sigma_i^2$  sono uguali tra loro), ha ancora funzione di distribuzione di  $\chi^2$  ma con  $\nu = n - 1$  gradi di libertà. In entrambi i casi il test d'ipotesi (compatibilità dei valori  $y_i$ ) viene eseguito come già descritto.

Supponiamo di avere a disposizione un campione di  $n$  misure  $y_i$ , con funzione di distribuzione normale e varianze  $\sigma_i^2$ , di una grandezza fisica  $Y$  ottenute in corrispondenza di  $n$  valori noti  $x_i$  di una grandezza fisica  $X$ . Usando questi dati, si vuole testare l'ipotesi che tra  $X$  e  $Y$  ci sia la relazione  $Y = f(X; a_1, \dots, a_K)$ , con  $a_1, \dots, a_K$  parametri noti. In questo caso l'ipotesi nulla è  $H_o: \mu_i = E[y_i] = f(x_i; a_1, \dots, a_K)$ ,  $i = 1, n$  e, se è vera, la statistica  $t$  di eq. (67) ha funzione di distribuzione di  $\chi^2$ . Il numero di gradi di libertà è  $\nu = n$  se i valori dei parametri  $a_j$  sono noti,  $\nu = n - k$  se sono stati ottenuti utilizzando il campione stesso per la loro stima. Noto  $\nu$ , si fissano  $\alpha$  e  $t_\alpha$ , si calcola  $t^*$  e, in caso, si rigetta l'ipotesi che tra  $X$  e  $Y$  ci sia la relazione  $Y = f(X; a_1, \dots, a_K)$ .

Altro caso importante di test di  $\chi^2$  è quello in cui si ha a disposizione un campione costituito da  $n$  valori  $x_i$  di una variabile casuale  $x$  che si ipotizza abbia funzione di distribuzione  $f(x)$ . Se  $n$  è sufficientemente grande, si può costruire l'istogramma corrispondente che consisterà di un numero  $N$  di intervalli da  $x_{min}$  a  $x_{max}$  in cui cadono tutti i valori  $x_i$ . Se si assume che i numeri di eventi negli intervalli dell'istogramma abbiano distribuzione multinomiale, essendo nota  $f(x)$  si possono calcolare valori di aspettazione, varianze e covarianze. Nel generico intervallo  $k$  centrato in  $x_k^*$  e di ampiezza  $\Delta_k$ , il valore di aspettazione del numero di eventi è  $\mu_k = E[n_k] = np_k$  e la varianza è  $\sigma_k^2 = np_k$ , dove  $p_k \simeq f(x_k^*)\Delta_k$ , se  $\Delta_k$  è sufficientemente piccolo. Se i valori  $\mu_k$  sono abbastanza grandi (tipicamente maggiori di 10) i numeri di eventi negli intervalli dell'istogramma hanno distribuzione molto simile alla distribuzione multinormale e la statistica

$$t(n_1, \dots, n_N) = (\vec{n} - \vec{\mu})^T V^{-1} (\vec{n} - \vec{\mu}), \quad (70)$$

o

$$t(n_1, \dots, n_N) = \sum_{k=1}^N \frac{(n_k - \mu_k)^2}{\mu_k} \quad (71)$$

se le covarianze sono trascurabili, ha funzione di distribuzione di  $\chi^2$ . Il numero di gradi di libertà è  $\nu = N - 1$  se i parametri che compaiono nella funzione  $f(x)$  sono noti: c'è comunque una relazione tra i dati dovuta alla normalizzazione. Se invece in  $f(x)$  compaiono  $k$  parametri stimati dal campione, il numero di gradi di libertà è  $\nu = N - 1 - k$ . Il test di  $\chi^2$  (con  $t$  come statistica di test) può quindi essere usato anche per eseguire il test d'ipotesi che la funzione di distribuzione di  $x$  sia  $f(x)$ . Questo esempio, e il prossimo, sono casi tipici di test di Pearson.

### 5.3 Test di indipendenza

Anche questo è un test d'ipotesi piuttosto diffuso, anche nel settore sanitario e in quello industriale, e permette di testare l'ipotesi che un certo numero di variabili casuali siano indipendenti, cioè che la funzione di distribuzione congiunta sia il prodotto delle funzioni di distribuzione delle singole variabili.

Nel caso di due variabili casuali  $x$  e  $y$ , avendo a disposizione un campione di  $n$  coppie di valori  $x_i, y_i$ , i dati vengono raggruppati in base a caratteristiche tipiche della popolazione (che possono anche essere l'appartenenza a certi intervalli di valori di  $x$ , e di  $y$ ). Si costruisce così la "tabella di contingenza", una matrice avente come elementi i numeri di eventi nei diversi gruppi i cui valori di aspettazione si possono calcolare nel caso di variabili indipendenti. Infine per il test d'ipotesi si utilizza il test di  $\chi^2$ .

\* per una descrizione più completa ed esempi vedi appunti

---

## 6 Appendice A - Distribuzione esponenziale o dei tempi di attesa

Se  $t_0 = 0$  è un istante arbitrario o il tempo al quale si verifica un evento e  $t$  è il tempo al quale si verifica l'evento successivo, allora  $t$  è una variabile casuale continua con funzione di distribuzione:

$$f(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

con  $\frac{1}{\tau} = \mu$  numero medio di eventi nell'unità di tempo.

\* Dimostrazione:

la probabilità che il primo evento successivo si verifichi in un piccolo intervallo di tempo  $(t, t + \delta t)$  è data dal prodotto (gli eventi sono indipendenti) delle probabilità di avere 0 eventi in  $(0, t)$ , cioè 0 eventi fino al tempo  $t$ , e di avere 1 evento in  $(t, t + \delta t)$ , cioè 1 evento in un intervallo  $\delta t$ :

$$P(t, t + \delta t) = P_0(t) \cdot P_1(\delta t).$$

Se l'intervallo  $\delta t$  è sufficientemente piccolo rispetto alla frequenza degli eventi, possiamo assumere che in  $\delta t$  non ci possano essere due eventi e che la probabilità di averne uno sia proporzionale a  $\delta t$  stesso, cioè che sia  $P_1(\delta t) = \mu \cdot \delta t$ . Per calcolare  $P_0(t)$ , consideriamo la probabilità di avere 0 eventi in  $0, t + \delta t$ :

$$P_0(t + \delta t) = P_0(t) \cdot P_0(t + \delta t) = P_0(t) \cdot (1 - \mu \cdot \delta t)$$

da cui

$$\frac{P_0(t + \delta t) - P_0(t)}{\delta t} = -\mu \cdot P_0(t).$$

Per  $\delta t$  che tende a zero, si ottiene l'equazione differenziale:

$$\frac{dP_0}{dt} = -\mu \cdot P_0$$

la cui soluzione è:

$$P_0(t) = A e^{-\mu t}.$$

Poiché deve essere  $P_0(0) = 1$ , si ha  $A = 1$ . Quindi  $P_0(t) = e^{-\mu t}$ . Notare che la stessa espressione si sarebbe potuta ottenere dalla distribuzione di Poisson, sostituendo  $\mu t$  a  $\nu$ . È quindi:

$$P(t, t + \delta t) = e^{-\mu t} \cdot \mu \delta t = f(t) \delta t$$

se il  $\delta t$  considerato è sufficientemente piccolo. La distribuzione degli intervalli di tempo tra un istante arbitrario e il primo evento successivo è perciò:

$$f(t) = \mu \cdot e^{-\mu t}.$$

Questa funzione di distribuzione, esponenziale negativa, è anche detta distribuzione dei tempi di attesa e si applica in tutti i fenomeni in cui gli eventi sono indipendenti, dal decadimento di particelle e nuclei radioattivi, alla rivelazione di particelle cariche nei raggi cosmici. Si può applicare anche al passaggio di automobili in un certo punto di una strada in condizioni di poco traffico regolare (senza semafori o simili) o ai terremoti in una certa regione.

Come si può facilmente verificare:

- è normalizzata:  $\int_0^{+\infty} \frac{1}{\tau} e^{-\frac{t}{\tau}} dt = 1$

- valore di aspettazione:  $E[t] = \int_0^{+\infty} t \cdot \frac{1}{\tau} e^{-\frac{t}{\tau}} dt = \tau$

- varianza:  $Var(t) = E[(t - \tau)^2] = \int_0^{+\infty} t^2 \cdot \frac{1}{\tau} e^{-\frac{t}{\tau}} dt - \tau^2 = \tau^2$

Il suo valore a  $t = 0$  è  $1/\tau$ . La retta tangente nell'origine interseca l'asse del tempo a  $t = \tau$  (detto vita media, nel caso dei decadimenti), valore al quale la funzione è diminuita di un fattore  $e$ .

Integrando la funzione, o usando la distribuzione cumulativa già ricavata, è immediato verificare che la probabilità che un evento accada in  $(\tau - \sigma_t, \tau + \sigma_t)$  è molto diversa da quella che si ha per una funzione di distribuzione di Gauss.

A questo punto si può facilmente calcolare la funzione di distribuzione del tempo di attesa del generico evento  $k$ . Seguendo il procedimento usato nel caso precedente si ha che la probabilità che il  $k$ -esimo evento si verifichi nell'intervallo di tempo  $(t, t + \delta t)$  è data dal prodotto della probabilità di avere  $k - 1$  eventi in  $(0, t)$  e della probabilità di avere 1 evento in  $(t, t + \delta t)$ :

$$P_k(t, t + \delta t) = P_{k-1}(t) \cdot P_1(\delta t).$$

Usando la distribuzione di Poisson, è  $P_{k-1}(t) = (\mu t)^{k-1} e^{-\mu t} / (k - 1)!$  e quindi

$$P_k(t, t + \delta t) = \frac{(\mu t)^{k-1} e^{-\mu t}}{(k - 1)!} \mu \delta t$$

e quindi la funzione di distribuzione cercata è

$$f_k(t) = \mu \frac{(\mu t)^{k-1} e^{-\mu t}}{(k - 1)!}.$$

È questo un tipo particolare di funzione di distribuzione Gamma, noto anche come funzione di distribuzione di Erlang di notevole uso pratico. Si può facilmente verificare che le funzioni  $f_k$  sono normalizzate e che:

$$E_k[t] = \frac{k}{\mu} = k\tau \quad \sigma_{t,k}^2 = \frac{k}{\mu^2} = k\tau^2.$$

Per  $k = 1$  si ottiene la distribuzione esponenziale; per  $k \geq 2$  le funzioni di distribuzione sono uguali a zero per  $t = 0$  e tendono ancora a zero per  $t \rightarrow \infty$ .

## 7 Appendice B - incertezze sistematiche e correlazioni

Questa appendice è dedicata a un esempio specifico di quanto introdotto in 1.3.4. Supponiamo di voler verificare che il periodo di oscillazione  $T$  di un pendolo semplice sia legato alla sua lunghezza  $l$  dalla relazione  $T = 2\pi\sqrt{l/g}$ .

Per fare ciò fissiamo  $n$  diverse lunghezze  $l_i$  e per ciascuna di queste misuriamo il corrispondente periodo  $T_i$ . Eseguendo le misure con un cronometro a comandi manuali, si ottengono i valori misurati  $T_i^m$  che sono caratterizzati da incertezze statistiche  $\sigma_{stat,i}$  (indicate anche con  $\sigma_{stat}$ ) che si possono stimare con una serie di misure ripetute. In assenza di errori sistematici sarà quindi

$$T_i^m = T_i^v + u_i$$

dove  $T_i^v$  è il valore vero (non noto) e  $u_i$  è l'errore accidentale di quella misura, anch'esso non noto, ma di cui si conosce la funzione di distribuzione  $N(0, \sigma_{stat,i}^2)$ .

Supponiamo ora che il cronometro utilizzato (lo stesso per tutte le misure) abbia una risposta caratterizzata da un offset  $x$ , cioè che un tempo  $x$ , positivo o negativo ma sempre uguale, venga aggiunto ad ogni misura di intervalli di tempo. Supponiamo di sapere, da misure eseguite su molti cronometri della stessa serie o da informazioni fornite dal costruttore, anche che  $x$  appartenga a una popolazione con distribuzione  $N(0, \sigma_{syst}^2)$  con  $\sigma_{syst}^2$  nota.

I periodi misurati nelle generiche misure  $i$  e  $j$  sono quindi

$$T_i^m = T_i^v + u_i + x, \quad T_j^m = T_j^v + u_j + x.$$

I valori di aspettazione dei periodi misurati sono i valori veri, infatti

$$E[T_i^m] = E[T_i^v] + E[u_i] + E[x] = T_i^v.$$

Usando la legge di propagazione della varianza, si ha anche

$$var(T_i^m) = var(u_i) + var(x) = \sigma_{stat,i}^2 + \sigma_{syst}^2.$$

La covarianza tra due diverse misure  $T_i^m$  e  $T_j^m$  può essere calcolata usando la relazione

$$cov(T_i^m, T_j^m) = E[T_i^m \cdot T_j^m] - E[T_i^m] \cdot E[T_j^m].$$

Si ottiene allora  $cov(T_i^m, T_j^m) = \sigma_{syst}^2$ , positiva e uguale per tutte le coppie di misure.

\* È infatti

$$\begin{aligned} E[T_i^m T_j^m] &= E[(T_i^v + u_i + x)(T_j^v + u_j + x)] = T_i^v T_j^v + E[u_i x] + E[x u_j] + E[x^2] \\ &= T_i^v T_j^v + cov(u_i x) + cov(x u_j) + var(x^2) = \sigma_{syst}^2 \end{aligned}$$

Il coefficiente di correlazione, infine, è:

$$\rho_{ij} = \frac{\sigma_{syst}^2}{\sqrt{(\sigma_{stat,i}^2 + \sigma_{syst}^2)(\sigma_{stat,j}^2 + \sigma_{syst}^2)}}.$$

Se  $\sigma_{stat,i(j)}^2 \gg \sigma_{syst}^2$ , cioè se le misure sono dominate dalle incertezze statistiche (indipendenti),  $\rho_{ij}$  ha un valore piccolo. Esso tende invece a 1 quando  $\sigma_{stat,i(j)}^2 \ll \sigma_{syst}^2$ , cioè quando l'incertezza dominante è quella sistematica, che è uguale per tutte le misure.

In generale, la correlazione tra le misure non può essere trascurata quando si vogliono utilizzare i dati per stimare i parametri della relazione tra le grandezze fisiche.

Un offset non è comunque l'unica possibile incertezza sistematica. Si potrebbe avere, ad esempio, un effetto “di scala”, tale che i valori misurati siano  $T_i^m = T_i^v + u_i + x \cdot T_i^v$ , dove  $x$  ha ora distribuzione  $N(1, \sigma_{syst}^2)$  con  $\sigma_{syst}^2$  nota. La complicazione è dovuta al fatto di avere incertezze sistematiche (e correlazioni) che dipendono dal valore vero, non noto, della grandezza fisica.

# Indice

<b>1</b>	<b>Teoria della probabilità</b>	<b>4</b>
1.1	Fenomeni statistici caratterizzati da una variabile casuale . . . . .	4
1.1.1	Variabili casuali discrete . . . . .	4
1.1.2	Variabili casuali continue . . . . .	5
1.1.3	Varianza e disuguaglianza di Chebyshev . . . . .	5
1.1.4	Momenti e funzioni generatrici dei momenti . . . . .	6
1.2	Alcune distribuzioni di probabilità e funzioni di distribuzione . . . . .	9
1.2.1	Funzione di distribuzione di Gauss . . . . .	9
1.2.2	Funzione di distribuzione uniforme . . . . .	11
1.2.3	Funzione di distribuzione esponenziale . . . . .	12
1.2.4	Distribuzione binomiale . . . . .	13
1.2.5	Distribuzione di Poisson . . . . .	14
1.2.6	Istogrammi . . . . .	16
1.2.7	Distribuzione Gamma . . . . .	17
1.2.8	Distribuzione di Erlang . . . . .	18
1.2.9	Distribuzione di $\chi^2$ . . . . .	20
1.3	Più variabili casuali . . . . .	23
1.3.1	Funzione di distribuzione congiunta, marginale e condizionata . . . . .	23
1.3.2	Valore di aspettazione e varianza . . . . .	24
1.3.3	Covarianza e coefficiente di correlazione . . . . .	25
1.3.4	Esempi di variabili casuali correlate . . . . .	26
1.3.5	Distribuzione multinomiale . . . . .	27
1.3.6	Distribuzione multinormale . . . . .	29
1.4	Funzioni di una variabile casuale . . . . .	34
1.5	Funzioni di più variabili casuali . . . . .	35
1.5.1	Somma di quadrati di variabili casuali . . . . .	37
1.5.2	Somma di variabili casuali . . . . .	38
1.5.3	Media aritmetica . . . . .	38
1.5.4	Somme di quadrati degli scarti . . . . .	39
<b>2</b>	<b>Inferenza statistica</b>	<b>42</b>
<b>3</b>	<b>Stima dei parametri</b>	<b>43</b>
3.1	Considerazioni generali . . . . .	43
3.2	Caratteristiche degli stimatori . . . . .	44
3.3	Metodo del Maximum Likelihood . . . . .	47
3.3.1	Stimatori . . . . .	47
3.3.2	Esempi . . . . .	48
3.3.3	Ulteriori considerazioni . . . . .	49
3.4	Metodo dei Minimi Quadrati . . . . .	51
3.4.1	Principio . . . . .	51
3.4.2	Il metodo dei Minimi Quadrati Lineare . . . . .	53

3.4.3	Incertezze statistiche . . . . .	55
<b>4</b>	<b>Intervalli di confidenza</b>	<b>57</b>
4.1	Misure con distribuzione di Gauss . . . . .	58
<b>5</b>	<b>Test d'ipotesi</b>	<b>60</b>
5.1	Test parametrici per variabili gaussiane . . . . .	61
5.2	Test di $\chi^2$ . . . . .	63
5.3	Test di indipendenza . . . . .	65
<b>6</b>	<b>Appendice A - Distribuzione esponenziale o dei tempi di attesa</b>	<b>67</b>
<b>7</b>	<b>Appendice B - incertezze sistematiche e correlazioni</b>	<b>70</b>