Note di probabilità e statistica per il corso di Laboratorio 2

a.a. 2015/16

Anna Martin
Dipartimento di Fisica, Università degli Studi di Trieste

versione preliminare, 22 gennaio 2016

Introduzione

Queste note, sintetiche e parziali, sono una raccolta di argomenti di probabilità e statistica trattati durante il corso di Labotatorio 2, ad uso degli studenti del corso. Non hanno alcuna pretesa di completezza e per collegamenti tra gli argomenti, dimostrazioni, ulteriori applicazioni e approfondimenti, si rimanda agli appunti presi durante il corso e ai molti testi / dispense esistenti [1].

Inoltre, queste note non includono il programma svolto nelle esercitazioni che, assieme agli esercizi e alle applicazioni visti in aula, è fondamentale per la comprensione degli argomenti trattati.

Non sono incluse neppure le nozioni di base di teoria della probabilità trattate nel corso di Laboratorio 1 (in particolare: definizioni di probabilità e Teorema di Bayes; definizioni di variabili casuali e distribuzioni /densità di probabilità, valore di aspettazione, varianza e deviazione standard; distribuzione uniforme e binomiale; funzione di distribuzione uniforme e di Gauss; errori accidentali, risoluzione di lettura e incertezze statistiche) usate durante il corso.

Infine, si ricorda che questa è ancora una versione preliminare delle note, con parti ancora da migliorare e soggetta a revisione periodica. A questo proposito, per migliorare le versioni future, chi legge è invitato a segnalare problemi ed errori.

[1] tra questi si segnalano, per le parti trattate nel corso:
- F. Bradamante, Note di analisi statistica dei dati, disponibili sul web
- A. Rotondi, P. Pedroni, A. Pievatolo, Probabilit Statistica e Simulazione, ed. Springer
- capitolo di statistica dal Particle Data Group, K.A. Olive et al. (Particle Data Group), Chin. Phys. C, 38, 090001 (2014) and 2015 update,
http://pdg.lbl.gov/2015/reviews/rpp2014-rev-statistics.pdf

1 Momenti

1.1 Momenti algebrici e centrali

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

Il momento algebrico (o semplice) di ordine k è definito come

$$\mu_k^* = E[x^k] = \int_{\Omega_x} x^k \cdot f(x)dx\,. \tag{1}$$

Si ha quindi che $\mu_0^* = 1$ per la condizione di normalizzazione della densità di probabilità $f(x)$, e che $\mu_1^* = \mu_x$, valore di aspettazione della variabile casuale.

Il momento centrale di ordine k è invece

$$\mu_k = E[(x-\mu_x)^k] = \int_{\Omega_x} (x-\mu_x)^k \cdot f(x)dx \tag{2}$$

ed è

- $\mu_0 = 1$ ancora per la condizione di normalizzazione della densità di probabilità,
- $\mu_1 = E[x] - \mu_x = 0$,
- $\mu_2 = \sigma_x^2$, varianza della variabile casuale.

Il momento μ_3 è invece legato all'asimmetria di $f(x)$ ed è $\mu_3 = 0$ per funzioni simmetriche rispetto a μ_x. Come indice di asimmetria (skewness) si usa spesso la quantità adimensionale $\gamma_1 = \mu_3/\mu_2^{3/2} = \mu_3/\sigma_x^3$. Il momento centrale di ordine 4 dà informazioni su quanto $f(x)$ è "piatta", e viene spesso introdotto il coefficiente di curtosi (kurtosis) $\gamma_2 = \mu_4/\mu_2^2 - 3$ che è 0 per una funzione di distribuzione di Gauss, -3 per una funzione di distribuzione uniforme.

In condizioni molto generali l'insieme dei momenti determina completamente la funzione di distribuzione. In particolare: se due funzioni di distribuzione $f_1(x)$ e $f_2(x)$ hanno le stesso insieme di momenti e se la loro differenza può essere sviluppata in serie di potenze, $f_1(x) = f_2(x)$.

* dimostrazione:
se $f_1(x) - f_2(x) = c_0 + c_1 x + c_2 x^2 + ...$ si ha

$$\begin{aligned}
\int_{\Omega_x} [f_1(x) \;-\; f_2(x)]^2 dx \;&=\; \int_{\Omega_x} [f_1(x) - f_2(x)]\,[c_0 + c_1 x + c_2 x^2 + ...]\,dx \;= \\
&= c_0 \int_{\Omega_x} [f_1(x) - f_2(x)]\,dx + c_1 \int_{\Omega_x} [f_1(x) - f_2(x)]\,x\,dx + c_2 \int_{\Omega_x} [f_1(x) - f_2(x)]^2\,x^2\,dx + ... \\
&= c_0[1-1] + c_1[\mu_1^* - \mu_1^*] + c_2[\mu_2^* - \mu_2^*] + ... \;=\; 0
\end{aligned}$$

e quindi $f_1(x) = f_2(x)$.
———-

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

$$\begin{aligned}
M_x^*(t) \;&=\; E[e^{tx}] \;=\; \int_{\Omega_x} e^{tx} f(x)\,dx \\
M_x(t) \;&=\; E[e^{t(x-\mu_x)}] \;=\; \int_{\Omega_x} e^{t(x-\mu_x)} f(x)\,dx
\end{aligned} \tag{3}$$

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

$$M_x(t) = e^{-t\mu_x} M_x^*(t). \tag{4}$$

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

* infatti

$$\begin{aligned} M_x^*(t) &= \int_{\Omega_x} [1 + xt + \frac{(xt)^2}{2!} + ...\frac{(xt)^n}{n!} + ...]\, f(x)\, dx \\ &= 1 + \mu_1^* t + \mu_2^* \frac{t^2}{2!} + ...\mu_n^* \frac{t^n}{n!} + ... \end{aligned}$$

e analoga per $M_x(t)$.

———-

Si ha quindi

$$\mu_k = \left[\frac{\partial^k M_x}{\partial t^k}\right]_{t=0}, \quad \mu_k^* = \left[\frac{\partial^k M_x^*}{\partial t^k}\right]_{t=0}. \tag{5}$$

Infine bisogna notare che, in base a quanto detto, se due funzioni di distribuzione hanno le stesse funzioni generatrici dei momenti, coincidono, proprietà che verrà utilizzata spesso nel seguito.

1.2 Esempi di funzioni generatrici dei momenti

In questo paragrafo sono ricavate le espressioni delle funzioni generatrici di alcune funzioni di distribuzione già studiate.

1.2.1 Distribuzione binomiale

La probabilità di avere k eventi favorevoli su n eventi indipendenti, se la probabilità che il singolo evento sia favorevole è p e la probabilità che sia sfavorevole è $q = 1 - p$, è data da

$$P(k; n, p) = \left(\begin{array}{c} n \\ k \end{array} \right) \cdot p^k \cdot q^{n-k}\,. \tag{6}$$

La funzione generatrice dei momenti algebrici è

$$M_x^*(t) = (pe^t + q)^n\,. \tag{7}$$

* dimostrazione:

$$\begin{aligned} M_k^*(t) &= E[e^{tk}] = \sum_{k=0}^{n} e^{tk} \left(\begin{array}{c} n \\ k \end{array} \right) \cdot p^k \cdot q^{n-k} \\ &= \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \cdot (pe^t)^k \cdot q^{n-k} = (pe^t + q)^n \end{aligned}$$

———-

Calcolando il valore della funzione e delle sue derivate in $t = 0$ si ottengono i risultati noti: $\mu_0^* = 1$ e $\mu_1^* = E[k] = np$.

Utilizzando eq. 4, si ottiene immediatamente

$$M_x(t) = e^{-t\,np} M_x^*(t) = (pe^{qt} + qe^{-pt})^n \,. \tag{8}$$

Calcolando i valori della funzione e delle sue derivate in $t = 0$ si trova $\mu_0 = 1$, $\mu_1 = 0$ e $\mu_2 = \sigma_k^2 = npq$, come previsto.

1.2.2 Funzione di distribuzione di Gauss

Per la funzione di distribuzione di Gauss

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

le funzioni generatrici dei momenti sono

$$M_x^*(t) = e^{\mu t + \sigma^2 t^2/2} \,, \quad M_x(t) = e^{\sigma^2 t^2/2} \tag{9}$$

* dimostrazione:
essendo

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

introducendo $y = x - \mu - t\sigma^2$ è

$$2tx\sigma^2 - (x-\mu)^2 = -(x - \mu - t\sigma^2)^2 + 2\mu t\sigma^2 + (t\sigma^2)^2 = -y^2 + [2\mu t + t^2\sigma^2]\sigma^2$$

e quindi

$$M_x^*(t) = e^{\mu t + t^2\sigma^2/2} \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{\frac{-y^2}{2\sigma^2}} dy = e^{\mu t + t^2\sigma^2/2} \,.$$

———-

Di nuovo, si può verificare che, calcolando i momenti dai valori delle funzioni e delle loro derivate in $t = 0$, si ottengono i risultati noti, cioè $\mu_0^* = 1\,,\ \mu_1^* = \mu\,,\ \mu_0 = 1\,,\ \mu_1 = 0\,,\ \mu_2 = \sigma^2$.

2 Alcune distribuzioni e densità di probabilità

Oltre alle già note distribuzioni e funzioni di distribuzione (uniforme, binomiale e di Gauss), ce ne sono altre particolarmente interessanti. Alcune di queste vengono intordotte in questo capitolo.

2.1 Distribuzione di Poisson

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

$$P_k = P(k;\nu) \quad = \quad \frac{\nu^k e^{-\nu}}{k!} \tag{10}$$

come si può vedere calcolando il limite della distribuzione binomiale.

* usando $p = \nu/n$ la distribuzione binomiale può essere riscritta nella forma

$$P(k) \quad = \quad \frac{n(n-1)\cdot(n-k+1)}{n^k}\frac{\nu^k}{k!}\left[\left(1-\frac{\nu}{n}\right)^{\frac{n}{\nu}}\right]^{\nu}\left(1-\frac{\nu}{n}\right)^{-k}$$

che al limite per $n \to \infty$ dà eq. (10).

———-

La distribuzione è normalizzata, e valore di aspettazione e varianza coincidono:

$$E[k] = \nu\,, \quad \sigma_k^2 = \nu.$$

* dimostrazione:

$$\begin{aligned}
\sum_{k=0}^{\infty} P_k &= e^{-\nu}\sum_{k=0}^{\infty}\frac{\nu^k}{k!} = 1 \\
E[k] &= \sum_{k=0}^{\infty} kP_k = e^{-\nu}\sum_{k=1}^{\infty}\frac{\nu^k}{(k-1)!} = e^{-\nu}\nu\sum_{r=0}^{\infty}\frac{\nu^r}{r!} = \nu \qquad (r = k-1) \\
E[k^2] &= \sum_{k=0}^{\infty} k^2 P_k = e^{-\nu}\nu\sum_{r=0}^{\infty}(1+r)\frac{\nu^r}{r!} = e^{-\nu}\nu(e^{\nu}+\nu e^{\nu}) = \nu+\nu^2 \\
\sigma_k^2 &= E[k^2] - (E[k])^2 = \nu.
\end{aligned}$$

———-

Notare che, se la deviazione standard aumenta con ν, l'incertezza relativa diminuisce come $1/\sqrt{\nu}$. La distribuzione ha una forma asimmetrica, ma tendo a diventare simmetrica all'aumentare di ν.

Le probabilità in funzione di k per $\nu = 2$, 5 e 10 sono visualizzate in fig. 1.

L'applicazione più tipica in fisica è relativa ai conteggi: supponiamo di avere in media ν eventi indipendenti nell'unità di tempo e quindi in media $\nu \cdot \Delta t$ eventi nell'intervallo di tempo Δt. La variabile casuale k, numero di eventi in Δt, ha distribuzione di Poisson con $E[k] = \nu \cdot \Delta t$

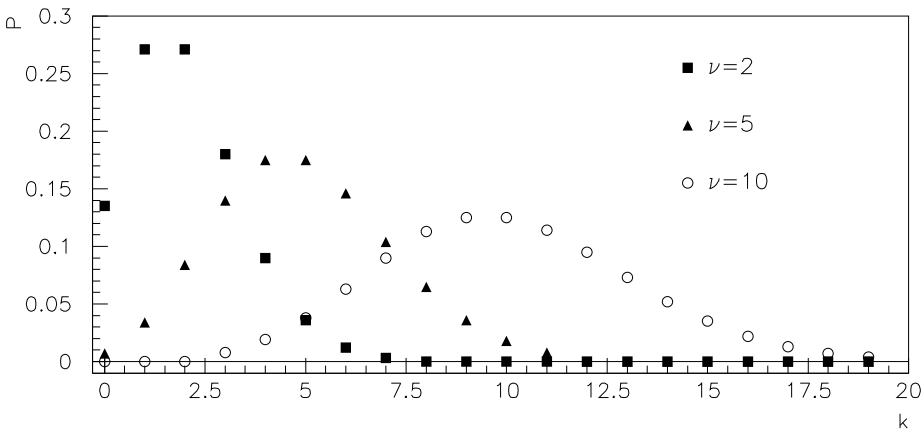

Figure 1: Distribuzione di Poisson per diversi valori di ν.

* Infatti possiamo dividere l'intervallo Δt in n intervalli di ampiezza $\delta t = \Delta t/n$ sufficientemente piccola, in modo che la probabilità di avere due o più eventi nello stesso intervallo sia 0. 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 $p = \nu \cdot \delta t$. Il valore di aspettazione di k è $E[k] = \nu \cdot \Delta t$. Al limite per $n \to \infty$ e $p \to 0$ si ottine la distribuzione di Poisson.

La funzione generatrice dei momenti algebrici è

$$M_k^*(t) = E[e^{kt}] = e^{\nu(e^t-1)} \tag{11}$$

* Infatti

$$E[e^{kt}] \;=\; \sum_{k=0}^{\infty} \frac{\nu^k e^{-\nu}}{k!} e^{kt} = e^{-\nu} \sum_{k=0}^{\infty} \frac{(\nu e^t)^k}{k!} = e^{-\nu} e^{\nu e^t} \;=\; e^{\nu(e^t-1)}$$

---

e la funzione generatrice dei momenti centrali è

$$M_k(t) = E[e^{t(k-\nu)}] = e^{-t\nu} M_k^*(t) = e^{\nu(e^t-1-t)}\,. \tag{12}$$

Da queste si riottiene:

$$M_k^*(t=0) \;=\; M_k(t=0) = 1 \tag{13}$$

$$\left[\frac{\partial M_x^*}{\partial dt}\right]_{t=0} \;=\; E[k] = \nu \tag{14}$$

$$\left[\frac{\partial M_x}{\partial dt}\right]_{t=0} \;=\; 0 \tag{15}$$

$$\left[\frac{\partial^2 M_x}{\partial dt^2}\right]_{t=0} \;=\; E[(k-\nu)^2] = \sigma^2 = \nu. \tag{16}$$

Inoltre, è

$$\left[\frac{\partial^3 M_x}{\partial dt^3}\right]_{t=0} = E[(k-\nu)^3] = \mu_3 = \nu \; ; \tag{17}$$

quindi il coefficiente di asimmetria è $\gamma_1 = \mu_3/\sigma^3 = 1/\sqrt{\nu}$ e tende a 0 (cioè la distribuzione diventa simmetrica) per $\nu \to \infty$. Di nuovo, la funzione diventa quasi simmetrica per ν abbastanza grande, dell'ordine di qualche decina.

Le funzioni generatrici si ottengono facilmente anche come limite delle funzioni generatrici dei momenti della distribuzione binomiale per $n \to \infty$ con $p \to 0$ e np costante.

* dimostrazione

———-

2.2 Distribuzione dei tempi di attesa

Per gli eventi descritti dalla distribuzione di Poisson è particolarmente interessante vedere qual'è la distribuzione dei tempi di attesa. Se $t = 0$ 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, t è una variabile casuale continua con funzione di distribuzione

$$f(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \tag{18}$$

con $1/\tau = \mu$, numero medio di eventi nell'unità di tempo.

* dimostrazione:
la probabilità che il primo evento successivo si verifichi in un piccolo intervello 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 δt:

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

Se l'intervallo δt è sufficientemente piccolo rispetto alla frequenza degli eventi, possiamo assumere che in δt non ci possano essere due eventi e che la probabilità di averne uno sia proporzionale a δt 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 δt che tende a zero, si ottiene l'equazione differenziale

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

la cui soluzione è

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

Poichè 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 in eq. (10) μt a ν. È quindi

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

se il δt 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, purchè non si considerino le scosse di assestamento correlate alle scosse principali.

Come si può facilmente verificare, la funzione è normalizzata e 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.

Il valore di aspettazione e la varianza sono

$$E[t] = \tau \ , \quad \sigma_t^2 = E[(t-\tau)^2] = \tau^2. \tag{19}$$

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

$$f_k(t) = \mu \frac{(\mu t)^{k-1} e^{-\mu t}}{(k-1)!}. \tag{20}$$

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 è

$$E_k[t] = \frac{k}{\mu} = \tau \quad \sigma_{t,k} = k\tau^2. \tag{21}$$

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

2.3 Distribuzione Gamma

La distribuzione Gamma una funzione di distribuzione che comprende, come casi particolari, anche le funzioni di distribuzione esponenziale, di Erlang e di χ^2. Viene utilizzata come modello generale dei tempi di attesa nella "teoria delle code".

Dipende da due paramentri, i numeri reali positivi α e β, e la sua espressione è

$$f(x;\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\frac{x}{\beta}} \tag{22}$$

con $x \geq 0$. La funzione $\Gamma(\alpha)$ è definita come

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

e valgono le relazioni

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)\,, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\,, \quad \Gamma(n) = (n-1)!$$

con n numero intero.

Usando la sostituzione $y = x/\beta$ e la definizione della funzione *Gamma* si può dimostrare che la funzione di distribuzione $f(x;\alpha,\beta)$ è nomalizzata e che è

$$E[x] = \alpha\beta \quad \sigma_x = \alpha\beta^2. \tag{23}$$

Infine la funzione generatrice dei momenti algebrici è

$$M_x^*(t) = (1-\beta t)^{-\alpha}. \tag{24}$$

La funzione di distribuzione di Erlang si ottiene per $\alpha = k,\ \beta = 1$, mentre la funzione di distribuzione di χ^2, descritta nel prossimo capitolo, si ottiene per $\alpha = \nu/2,\ \beta = 2$ con ν intero maggiore di zero.

2.4 Distribuzione di χ^2

La funzione di distribuzione di χ^2 è molto diffusa ed è la funzione di distribuzione di variabili casuali che sono somme dei quadrati di variabili casuali con funzione di distribuzione normale standard, come verrà dimostrato più avanti. Dipende da un solo parametro ($\nu \geq 1$), detto numero di gradi di libertà che, nel caso della somma dei quadrati di variabili casuali, coincide con il numero di addendi indipendenti.

La sua espressione è

$$f(z;\nu) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\, e^{-\frac{z}{2}}\, z^{\frac{\nu}{2}-1} \tag{25}$$

con $z \geq 0$. Anche in questo caso è facile dimostrare che la funzione è normalizzata, e valore di aspettazione e varianza sono

$$E[z] = \mu_z = \nu\,, \quad \sigma_z^2 = 2\nu = 2\mu_z. \tag{26}$$

Le funzioni generatrici dei momenti algebrici e centrali sono

$$M_z^*(t) = (1-2t)^{-\frac{\nu}{2}}\,, \quad M_z(t) = e^{-t\nu}(1-2t)^{-\frac{\nu}{2}}. \tag{27}$$

* Si ha

$$E[e^{tz}] = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty e^{-\frac{z}{2}(1-2t)}\, z^{\frac{\nu}{2}-1}\, dz$$
$$= \frac{1}{\Gamma\left(\frac{\nu}{2}\right)(1-2t)^{\frac{\nu}{2}}}\Gamma\left(\frac{\nu}{2}\right) = (1-2t)^{-\frac{\nu}{2}}$$

avendo fatto la sostituzione $y = z(1-2t)/2$.

———-

Dalle funzioni generatrici dei momenti si riottengono valore di aspettazione e varianza, e dalla derivata terza di $M_z(t)$ si ottiene il momento centrale di ordine 3 $\mu_3 = 8\nu$. Il coefficiente di asimmetria è quindi $\gamma_1 \sim 1/\sqrt{\nu}$, che tende a 0 per $\nu \to \infty$.

Come è facile calcolare, e come si vede da fig. 2, la funzione diverge a $z = 0$ per $\nu = 1$, è la funzione esponenziale $\frac{1}{2}e^{-z/2}$ per $\nu = 2$, ha un andamento $\sim \sqrt{z}e^{-z/2}$ per $\nu = 3$ ...

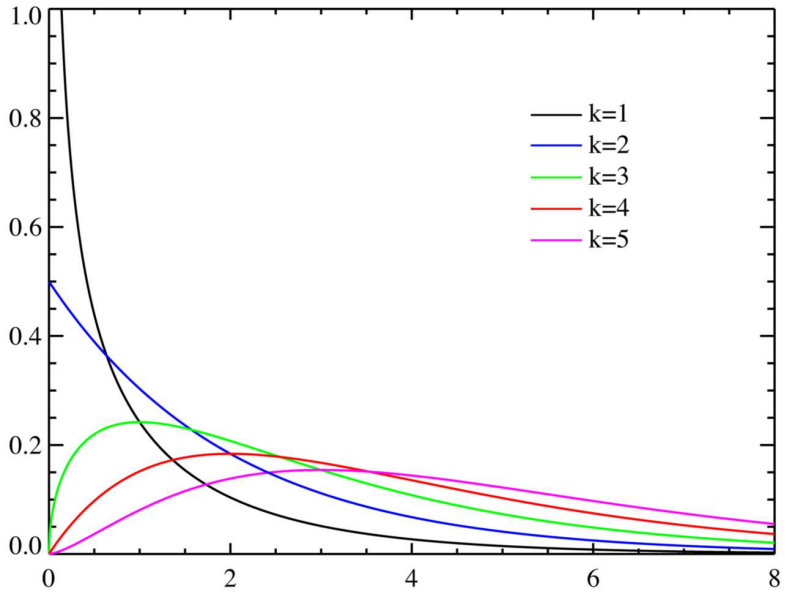

Figure 2: Funzione di distribuazione di χ^2 per diversi valori di ν.

Come vedremo, la funzione cumulativa è molto usata: si trova tabulata in quasi tutti i libri di statistica ed esistono diversi "calcolatori" disponibili sul web. Nelle tavole vengono generalmente dati i valori di z_α/ν per i quali $P(z/\nu > z_\alpha/\nu) > \alpha$. In genere si trovano solo valori per ν relativamente piccoli, in quanto, se ν è grande, z/ν ha una distribuzione molto simile a una distribuzione di Gauss con valore di aspettazione 1 e varianza $2/\nu$.

2.5 Distribuzione di Cauchy

La funzione di distribuzione di Cauchy (o Breit-Wigner, o Lorentz) ha la forma

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad -\infty \le x \le +\infty. \tag{28}$$

È una funzione simmetrica rispetto a $x = 0$, ma nonostante ciò il valore di aspettazione di x rigorosamente non è definito. Tutti i momenti divergono e la varianza non è finita (quindi in particolare non si può utilizzare neppure la diseguaglianza di Chebichev). Il valore della funzione a $x = 0$ è $1/\pi$ e i valori $1/2\pi$ si hanno per $x = \pm 1$. La larghezza a metà altezza è quindi 2.

La funzione di distribuzione di Cauchy descrive la distribuzione in massa invariante delle particelle prodotte nel decadimento di particelle instabili (o risonanze). La distribuzione in massa invariante M è infatti descritta dalla funzione

$$f(M) = \frac{1}{\pi} \frac{\Gamma/2}{(\Gamma/2)^2 + (M - M_0)^2} \tag{29}$$

che si ottiene da eq. (28) con il cambiamento di variabile

$$x = \frac{M - M_0}{\Gamma/2}$$

dove M_0 è la massa della particella che decade. La funzione è ora simmetrica rispetto a M_0 ed ha un'ampiezza a metà altezza pari a Γ.

3 Più variabili casuali

La parte introduttiva su: *funzione di distribuzione congiunta; valore di aspettazione e varianza; funzione di distribuzione marginale; funzione di distribuzione condizionata; variabili casuali indipendenti e non* non è inclusa (vedi Laboratorio 1 e appunti).

3.1 Covarianza

Un parametro importante nel caso in cui più variabili casuali intervengano nel fenomeno è la covarianza tra due variabili.

Nel caso di due sole variabili x e y, con densità di probabilità congiunta $f(x,y)$, valore di aspettazione e varianza di x sono dati da

$$E[x] \;=\; \mu_x = \int_{\Omega_x}\int_{\Omega_y} x \cdot f(x,y)\,dxdy \tag{30}$$

$$\begin{aligned} var(x) &= \sigma_x^2 = E[(x-\mu_x)^2] = \int_{\Omega_x}\int_{\Omega_y} (x-\mu_x)^2 \cdot f(x,y)\,dxdy \\ &= E[x^2] - \mu_x^2 = \int_{\Omega_x}\int_{\Omega_y} x^2 \cdot f(x,y)\,dxdy \;-\; \mu_x^2 \end{aligned} \tag{31}$$

e formule analoghe per y. La covarianza è invece definita come

$$\begin{aligned} cov(x,y) &= E[(x-\mu_x)(y-\mu_y)] = \int_{\Omega_x}\int_{\Omega_y} (x-\mu_x)(y-\mu_y) \cdot f(x,y)\,dxdy \\ &= E[xy] - \mu_x\mu_y\,. \end{aligned} \tag{32}$$

La covarianza è un numero, positivo o negativo, ma in genere diverso da 0. È uguale a zero per variabili indipendenti, quando cioè la densità di probabilità congiunta è il prodotto delle densità di probabilità di x e di y.

Per la diseguagkianza di Cauchy-Swarz deve essere

$$|cov(x,y)| \le \sigma_x\sigma_y \tag{33}$$

dove l'uguaglianza si ha nel caso di relazione lineare tra x e y.

* dimostrazione

———-

Il coefficiente di correlazione, definito come

$$\rho_{xy} = \frac{cov(x,y)}{\sigma_x\sigma_y} \tag{34}$$

deve quindi essere sempre minore o uguale a 1 in valore assoluto: $-1 \le \rho_{xy} \le +1$. Se $\rho_{xy} = \pm 1$ le variabili casuali x e y sono completamente correlate, positivamente o negativamente; è questo il caso di variabili legate da una relazione lineare. Se $\rho_{xy} = 0$ le variabili sono scorrelate. Questa e' una condizione necessaria ma non sufficiente perchè le variabili

siano indipendenti: se le variabili sono indipendi, la loro covarianza è zero, ma il viceversa non è sempre vero.

Nel caso di n variabili casuali x_i definite in Ω_{x_i}, con funzione di distribuzione congiunta $f(x_1, ..., x_n)$ e valori di aspettazione μ_i, $i = 1, n$, è

$$cov(x_i, x_j) = \int_{\Omega_{vec_x}} x_i x_j f(x_1, ..., x_n) \cdot dx_1 \cdot ... \cdot dx_n - \mu_i \mu_j \tag{35}$$

e il coefficiente di correlazione è ancora

$$\rho_{ij} = \frac{cov(x_i, x_j)}{\sigma_i \sigma_j} \tag{36}$$

con σ_i deviazione standard di x_i.

In genere si introduce la matrice delle covarianze con elementi $V_{ij} = \rho_{ij}\sigma_i\sigma_j$:

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{21}\sigma_2\sigma_1 & \cdots & \rho_{n1}\sigma_n\sigma_1 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{n2}\sigma_n\sigma_2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix} \tag{37}$$

che è una matrice simmetrica, diagonale se le variabili sono indipendenti. Come vedremo, la matrice delle covarianze permette di semplificare notevolmente la notazione nel caso di più variabili casuali.

3.2 Funzioni generatrici dei momenti

Nel caso di più variabili casuali si introducono le funzioni generatrici dei momenti "congiunte". Date n variabili casuali $x_1, ... x_n$ con funzione di distribuzione $f(x_1, ...x_n)$, valori di aspettazione $E[x_i] = \mu_i$ e varianze σ_i^2, la funzione generatrice dei momenti algebrici è una funzione di n variabili $t_1, ...t_n$ definita da

$$M^*(t_1, ...t_n) \quad = \quad E\left[e^{\sum_{i=1}^n t_i x_i}\right], \tag{38}$$

che, nel caso di variabili indipendenti è semplicemente il prodotto delle singole funzioni generatrici:

$$M^*(t_1, ...t_n) \quad = \quad M_x^*(t_1) \cdot M_x^*(t_2) ... M_x^*(t_n)\,. \tag{39}$$

Si verifica facilmente che è

$$[M^*(t_1, ...t_n)]_{\vec{t}=0} = 1 \qquad \mu_{1,i}^* = \left[\frac{\partial M_{x_i}^*}{\partial t_i}\right]_{\vec{t}=0} = \mu_i\,.$$

La funzione generatrice dei momenti centrali è

$$M(t_1, ...t_n) \quad = \quad E\left[e^{\sum_{i=1}^n t_i (x_i - \mu_i)}\right] = e^{-\sum_{i=1}^n t_i \mu_i} M^*(t_1, ...t_n)\,. \tag{40}$$

Si ha $[M(t_1,...t_n)]_{\vec{t}=0} = 1$ e

$$\begin{aligned}
\frac{\partial M_{x_i}}{\partial t_i} &= \int_{\Omega_{\vec{x}}} (x_i-\mu_i) e^{\sum_i t_i(x_i-\mu_i)} f(x_1,...x_n) dx_1...dx_n\,, \quad i.e. \left[\frac{\partial M_{x_i}}{\partial t_i}\right]_{\vec{t}=0} = 0 \\
\frac{\partial^2 M_{x_i}}{\partial t_i^2} &= \int_{\Omega_{\vec{x}}} (x_i-\mu_i)^2 e^{\sum_i t_i(x_i-\mu_i)} f(x_1,...x_n) dx_1...dx_n\,, \quad i.e. \left[\frac{\partial^2 M_{x_i}}{\partial t_i^2}\right]_{\vec{t}=0} = \sigma_i^2 \\
\frac{\partial^2 M_{x_i}}{\partial t_i \partial t_j} &= \int_{\Omega_{\vec{x}}} (x_i-\mu_i)(x_j-\mu_j) e^{\sum_i t_i(x_i-\mu_i)} f(x_1,...x_n) dx_1...dx_n\,, \\
& \quad i.e. \left[\frac{\partial^2 M_{x_i}}{\partial t_i \partial t_j}\right]_{\vec{t}=0} = cov(x_i,x_j)\,.
\end{aligned}$$

3.3 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.

Un altro esempio importante è costituito da misure di grandezze fisiche affette da incertezze sistematiche, caso trattato in questo paragrafo. Per chiarezza, si ricorda che per incertezza sistematica (non 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.

Supponiamo di voler verificare che il periodo di oscillazione T di un pendolo semplice è 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 i valori misurati T_i^m sono caratterizzati da incertezze statistiche $\sigma_{stat,i}$ (o σ_{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, pure non noto, ma di cui si conosce la funzione di distribuzione $N(0,\sigma^2_{stat,i})$.

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 lo stesso, venga aggiunto ad ogni misura di intervalli di tempo. Supponiamo anche che, da misure eseguite su molti cronometri della stessa serie o da informazioni formite dal costruttore, si sappia che x appartenga a una popolazione con distribuzione $N(0,\sigma^2_{syst})$ con σ^2_{syst} nota. I periodi misurati nelle generiche misure i e j sono quindi

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

I valori di aspettazione dei periodi misurati sono i valori veri, infatti è $E[T_i^m] = E[T_i^v] + E[u_i] + E[x] = T_i^v$. Usando la legge di propagazione della varianza, si ha anche $var(T_i^m) = var(u_i) + var(x) = \sigma^2_{stat,i} + \sigma^2_{syst}$.

La covarizanza tra due diverse misure T_i^m e T_j^m può essere calcolata usando la relazione $cov(T_i^m, T_j^m) = E[T_i^m T_j^m] - E[T_i^m]E[T_j^m]$. Si ottiene $cov(T_i^m, T_j^m) = \sigma^2_{syst}$, positiva e uguale per tutte le coppie di misure.

* È infatti

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

———-

Il coefficiente di correlazione, infine, è':

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

Se $\sigma^2_{stat,i(j)} >> \sigma^2_{syst}$, cioè se le misure sono dominate dalle incertezze statistiche (indipendenti), ρ_{ij} è piccolo. Tende invece a 1 quando $\sigma^2_{stat,i(j)} << \sigma^2_{syst}$ cioè quando l'incertezza dominante è quella sistematica, uguale per tutte le misure. Risultati molto ragionevoli. La correlazione tra le misure non può essere in generale trascurata quando si vogliano 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 $T_i^m = T_i^v + u_i + xT_i^v$, dove x ha ora distribuzione $N(1, \sigma^2_{syst})$ con σ^2_{syst} nota. La complicazione è dovuta al fatto di avere incertezze sistematiche (e correlazioni) che dipendono dal valore vero, non noto, della grandezza fisica.

3.4 Alcune distribuzioni congiunte

Qui vengono trattate solo due distribuzioni particolarmente utili per gli scopi del corso: la distribuzione di probabilità multinomiale (variabili casuali discrete) e la distribuzione multinormale (variabili casuali continue).

3.4.1 Distribuzione multinomiale

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 eq. (6).

Se le alternative sono $m > 2$, con probabilità $p_1, ...p_m$ che il singolo evento si verifichi secondo le diverse modalità (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

$$P(k_1, ...k_m) = \frac{n!}{k_1! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m}. \tag{41}$$

La dimostrazione è simile a quella del caso della distribuzione binomiale. Ora il numero delle possibili sequenze è $n!/(k_1! \cdots k_m!)$.

Valori di aspettazione e varianze sono

$$E[k_i] = np_i \qquad \sigma^2_{k_i} = np_i(1-p_i) \tag{42}$$

come si può vedere abbastanza facilmente. Per giustificare queste relazioni, 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, le variabili k_i sono correlate, ed è

$$cov(k_i, k_j) = E[(k_i - np_i)(k_j - np_j)] = -np_ip_j, \tag{43}$$

diversa da zero e negativa.

* Dimostrazione, nel caso di tre sole modalità:

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

$$\begin{aligned} E[k_1k_2] &= n(n-1)p_1p_2 \sum_{k_1,k_2=1,n} \frac{(n-2)!}{(k_1-1)!(k_2-1)!(n-k_1-k_2)!}p_1^{k_1-1}p_2^{k_2-1}(1-p_1-p_2)^{n-k_1-k_2} \\ &= n(n-1)p_1p_2 \end{aligned}$$

$$cov(k_1,k_2) = E[k_1k_2] - E[k_1]E[k_2] = n^2p_1p_2 - np_1p_2 + n^2p_1p_2 = -np_1p_2$$

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. Se però il numero di intervalli con eventi è elevato, come spesso accade, è $p_i << 1$ per ogni i e si ottiene: $E[k_i] = np_i$, $\sigma^2_{k_i} \simeq np_i$, 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 massimo è circa -0.1, 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 ν (ad esempio, invece di contare n=100 eventi, contiamo gli eventi in un intervallo di tempo in cui in media ce ne sono ν=100):

$$P(n) = \frac{e^{-\nu}}{n!}\nu^n\,. \tag{44}$$

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

$$\begin{aligned} P(n,k_1,...k_m) &= \frac{e^{-\nu}}{n!}\nu^n\frac{n!}{k_1!\cdots k_m!}p_1^{k_1}\cdots p_m^{k_m} \\ &= \frac{e^{-\nu p_1}}{k_1!}(\nu p_1)^{k_1}\cdots\frac{e^{-\nu p_m}}{k_m!}(\nu p_m)^{k_m} \end{aligned} \tag{45}$$

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.

3.4.2 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 variabili 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:

$$f(x_1, x_2) = f_1(x_1) f_2(x_2) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-Q^2/2} \tag{46}$$

dove

$$Q^2 = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}. \tag{47}$$

Le funzioni di distribuzione marginali e condizionate sono

$$f_{x_1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 = f_1(x_1) \qquad f_{x_2}(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_1 = f_2(x_2) \tag{48}$$

$$f(x_1|x_2^*) = \frac{f(x_1, x_2^*)}{f_{x_2}(x_2^*)} = f_1(x_1) \qquad f(x_2|x_1^*) = \frac{f(x_1^*, x_2)}{f_{x_1}(x_1^*)} = f_2(x_2) \tag{49}$$

$$\tag{50}$$

e corrispondono rispettivamente alle intersezioni (riscalate) della superficie $f(x_1, x_2)$ con piani paralleli al piano x_2, z o x_1, z, e alle proiezioni sul piano x_1, z o x_2, z.

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 (μ_1, μ_2) di equazione

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = c\,. \tag{51}$$

Per $c = 1$ l'ellisse ha semiassi σ_1 e σ_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, come vedremo, Q^2 è una variabile con funzione di distribuzione di χ^2 a 2 gradi di libertà e la probabilità che sia $Q^2 < 1$ è la probabilità di avere $\chi_2^2 < 1$. Per $c = 4$ l'ellisse ha semiassi $2\sigma_x$ e $2\sigma_y$, 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 $f_i(x_i) = N(\mu_i, \sigma_i^2)$, la funzione di distribuzione congiunta è ancora data dal prodotto delle funzioni di distribuzione:

$$f(x_1, ..x_n) = \frac{1}{(2\pi)^{n/2}\sigma_1 ... \sigma_n} e^{-Q^2/2} \quad \text{con } Q^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2}. \tag{52}$$

Usando la matrice delle covarianze

$$V = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \tag{53}$$

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

$$f(\vec{x}) = \frac{1}{(2\pi)^{n/2}|detV|^{1/2}} e^{-Q^2/2} \quad \text{con } Q^2 = (\vec{x}-\vec{\mu})^T V^{-1} (\vec{x}-\vec{\mu}) \tag{54}$$

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

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{21}\sigma_2\sigma_1 & \cdots & \rho_{n1}\sigma_n\sigma_1 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{n2}\sigma_n\sigma_2 \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \rho_{2n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{pmatrix} \tag{55}$$

L'espressione esplicita della funzione di distribuzione multinormale si può ottenere facilmente nel caso di due sole variabili x_1 e x_2 con valori di aspettazione μ_1 e μ_2, varianze σ_1^2 e σ_2^2, e coefficiente di correlazione ρ_{12}. In questo caso la matrice delle covarianze diventa

$$V = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{56}$$

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

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} e^{-Q^2/2} \tag{57}$$

con

$$Q^2 = \frac{1}{1-\rho_{12}^2}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho_{12}\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]. \tag{58}$$

Per $Q^2 = c$ si ottiene ancora un'ellisse, ancora con centro (μ_1, μ_2) ma questa volta non riferita agli assi principali; l'angolo di rotazione ϕ può essere maggiore o minore di $\pi/2$, in accordo con il segno del coefficiente di correlazione.

L'ellisse $Q^2 = 1$ è rappresentata schematicamente in fig. 3. 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 $(\mu_2 \pm \sqrt{1-\rho_{12}^2}\sigma_2)$;
- la retta $x_2 = \mu_2$ interseca l'ellisse nei punti di coordinate $(\mu_1 \pm \sqrt{1-\rho_{12}^2}\sigma_1)$;

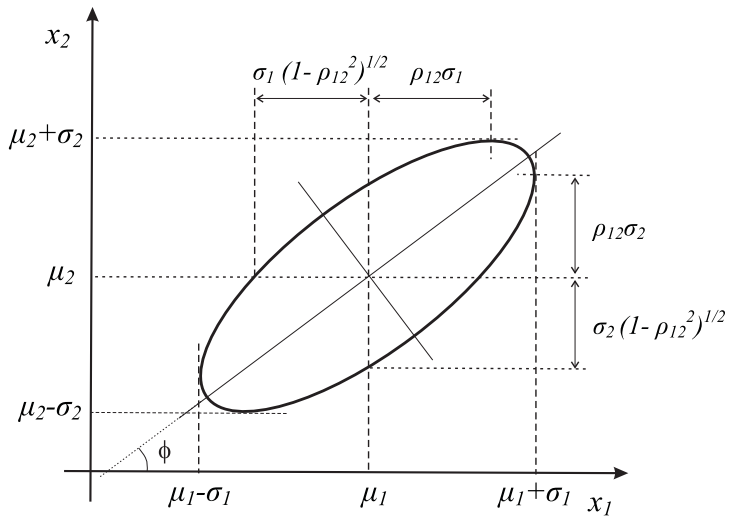

Figure 3: Ellisse $Q^2 = 1$.

- le 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 rette tangenti parallele all'asse x_1 sono $x_2 = \mu_2 \pm \sigma_2$ e i punti di tangenza hanno coordinate $(\mu_2 \pm \sigma_2, \mu_1 \pm \rho_{12}\sigma_1)$.

Queste relzioni permettono di determinare varianze e covarianza 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.

Le funzioni di distribuzione marginali sono le distribuzioni di Gauss delle singole variabili:

$$f_{x_1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2) dx_2 = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \tag{59}$$

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

* dimostrazione

———-

Si può ricavare facilmente che la funzione di distribuzione condizionata per x_2 (e per x_1, scambiando gli indici 1 e 2) è

$$f(x_2|x_1^*) = \frac{f(x_2|x_1^*)}{f_{x_1}(x_1^*)} = \frac{1}{\sqrt{2\pi}\sigma'_2} e^{-\frac{(x_2-\mu'_2)^2}{2\sigma'^2_2}} \tag{60}$$

con

$$\sigma'_2 = \sigma_2\sqrt{(1-\rho_{12})^2}\,, \qquad \mu'_2 = \mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(x_1^* - \mu_1)\,. \tag{61}$$

La varianza è quindi costante e ridotta del fattore $(1-\rho_{12})$, mentre il valore di aspettazione di x_2 dipende da x_1^* La retta $x_2 = \mu_2 + \rho_{12}\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$ è il luogo dei punti di massimo della funzione di distribuzione condizionata, ed è detta retta di regressione. Coincide con la retta che congiunge i due punti di tangenza $(\mu_1 \pm \sigma_1, \mu_2 \pm \rho_{12}\sigma_2)$.

Facendo riferimento a quanto visto per le funzioni generatrici dei momenti per più variabili casuali, si possono ricavare le funzioni generatrici dei momenti e verificare che le derivate seconde forniscono effettivamente varianze e covarianza. Nel caso di variabili casuali indipendenti è

$$M_{\vec{x}}(\vec{t}) = e^{\sum_i \sigma_i^2 t_i^2/2}, \qquad M^*_{\vec{x}}(\vec{t}) = e^{\sum_i (\sigma_i^2 t_i^2 + 2\mu_i t_i)/2} \tag{62}$$

mentre se le covarianze sono diverse da zero è

$$M_{\vec{x}}(\vec{t}) = e^{\frac{1}{2}\vec{t}^T V \vec{t}} \tag{63}$$

Per gli esempi di variabili con distribuzione multinormale si rimanda agli appunti.

4 Funzioni di variabili casuali

4.1 Funzioni di una o più variabili casuali

In generale, se $y = y(x)$ è una funzione della variabile casuale x con funzione di distribuzione $f(x)$, y è anche una variabile casuale con funzione di distribuzione $g(y)$ data da

$$g(y) = f[x(y)] \left| \frac{dx(y)}{dy} \right|. \tag{64}$$

nel caso di corrispondenza biunivoca (e, ovviamente, se la funzione $y = y(x)$ si può invertire).

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

$$g(y) = \sum_{j=1}^{m} f[x_j(y)] \left| \frac{dx_j(y)}{dy} \right|. \tag{65}$$

Alcuni esempi di applicazione della relazione (64) sono i seguenti:

- 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 $N(0, 1)$, $y = x^2$ ha funzione di distribuzione di χ^2 a un grado di libertà.

Spesso comunque non è necessario calcolare la funzione di distribuzione di y ma è sufficiente calcolarne valore di asptettazione e varianza con le formule approssimate

$$\mu_y \;=\; E[y] \;\simeq\; y(x = \mu_x) + \frac{1}{2} \left[\frac{d^2 y}{dx^2} \right]_{x=\mu_x} \cdot \sigma_x^2 \;\simeq\; y(x = \mu_x), \tag{66}$$

$$\sigma_y^2 \;=\; var(y) \;\simeq\; \left(\left[\frac{dy}{dx} \right]_{x=\mu_x} \right)^2 \cdot \sigma_x^2 \tag{67}$$

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 $x = \mu_x$.

Nel caso di funzione di più variabili casuali, $y = y(x_1, x_2, ..., x_n)$ con x_i n variabili casuali di cui sono noti valori di aspettazione e varianze, si può ancora utilizzare lo sviluppo in serie di $y(x_1, x_2, ..., x_n)$ per scrivere le formule approssimate per valore aspettazione e varianza di y (spesso sufficienti, come per le misure indirette):

$$\mu_y = E[y] \simeq y(\mu_{x_1}, \mu_{x_2}, ..., \mu_{x_n}), \tag{68}$$

$$\begin{aligned} \sigma_y^2 \;&\simeq\; \sum_{k=1}^{n} \left(\left[\frac{\partial y}{\partial x_k} \right]_{\vec{x}=\vec{\mu}_x} \right)^2 \sigma_{x_k}^2 \;+ \\ &+\; \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left[\frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}_x} \left[\frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}_x} cov(x_i, x_j). \end{aligned} \tag{69}$$

in cui, per variabile x_i correlate, i termini contenenti le covarianze non possono essere trascurati e compaiono moltiplicati per le derivate parziali (con segno) di y.

Per determinare la funzione di distribuzione di $y = y(x_1, x_2, ..., x_n)$ con x_i variabili casuali con funzione di distribuzione congiunta $f(x_1, x_2, ..., x_n)$, nel caso generale si introducono n variabili casuali y_i

$$\begin{aligned} y_1 &= y_1(x_1, x_2, ..., x_n) \\ y_2 &= y_2(x_1, x_2, ..., x_n) \\ &... \\ y_n &= y_n(x_1, x_2, ..., x_n) \end{aligned} \tag{70}$$

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

$$g(y_1, y_2, ..., y_n) = f[x_1(y_1, y_2, ..., y_n), x_2(y_1, y_2, ..., y_n), ..., x_n(y_1, y_2, ..., y_n)] \cdot |J| \tag{71}$$

dove

$$J = \left| \frac{\partial(x_1, x_2, ..., x_n)}{\partial(y_1, y_2, ..., y_n)} \right| \tag{72}$$

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

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 $g(y_1, y_2, ..., y_n)$, integrandola su tutte le variabili $y_j \neq y_1$, si ottiene la funzione di distribuzione marginale di y_1.

Come esercizio, per rendersi conto della difficoltà del procedimento, verificare che la somma (e la differenza) di due variabili casuali x_1, x_2 con distribuzione uniforme ha funzione di distribuzione triangolare.

Come vedremo nel seguito, in alcuni casi importanti è possibile però ottenere la funzione di distribuzione di una funzione di variabili casuali con metodi più semplici, come l'uso delle funzioni generatrici dei momenti.

In questo capitolo verranno considerrate solo alcune funzioni di varibili casuali. Per capirne l'importanza è opportuno introdurre prima il concetto di campione.

4.2 Popolazione, campione e funzioni di variabili casuali

Una variabile cauale x, in fisica, è associata alle misure relative a un ben definito fenomeno, o meglio, di una certa grandeza fisica. La funzione di distribuzione $f(x)$ riassume il risultato di tutte le possibili misure, eseguite nelle stesse condizioni, e ripetute un numero

infinito di volte: descrive quella che viene chiamata la popolazione (o universo). Se nota, permette di calcolare la probabilità di ottenere un valore misurato in un qualsiasi intervallo dei possibili valori di x. Essendo impossibile eseguire un numero infinito di misure, il concetto di popolazione è quindi una idealizzazione la cui conoscenza completa non può in pratica essere raggiunta. Un esperimento reale consiste in un numero finito di misure e una sequenza di n misure x_1, $x_2, ... x_n$ della stessa grandezze fisica costituisce un campione di dimensione n. Il campione è quindi un sottoinsieme della popolazione. Poichè ripetendo la misura altre n volte si otterranno risultati diversi, cioè un campione diverso, il campione è costituito da variabili casuali ed è anche detto campione casuale o aleatorio. Nel seguito si cercherà di indicare il campione casuale con lettere maiuscole mentre il campione specifico verrà indicato con lettere minuscole, anche se dal contesto il significato dovrebbe essere chiaro.

Lo scopo dell'inferenza statistica è proprio quello di fornire i metodi per ottenere, a partire da un campione finito, informazioni sull'intera popolazione, tipicamente stimando i parametri (incogniti) che compaiono nella funzione di distribuzione o eseguendo test sul campione a disposizione. Ciò viene fatto ricorrendo a funzioni del campione cioè a funzioni di variabili casuali. Un campione di dimensione n può essere infatti pensato come l'insieme di n valori assunti da una variabile casuale, X ma anche come l'insieme dei valori assunti da n variabili casuali X_1, $X_2, ... X_n$, non sempre indipendenti, e tutte con la stessa funzione di distribuzione $f(X)$. Le funzioni del campione sono quindi funzioni delle variabili casuali X_i.

4.3 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

$$Z = \sum_{i=1}^{n} X_i^2. \tag{73}$$

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

* Infatti, come visto nel paragrafo 4.1, il quadrato di una variabile casuale X con distribuzione $N(0,1)$ è una variabile casuale con funzione di distibuzione di χ^2_1 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

$$M^*_Z(t) = E[e^{t\sum_{i=1}^{n} X_i^2}] = \prod_i E[e^{tX_i^2}] = (1-2t)^{-n/2} \tag{74}$$

che è la funzione generatrice dei momenti algebrici della funzione di distribuzione di χ^2_n. ———-

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

$$Z = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2} \tag{75}$$

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

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

Inoltre a questo punto è chiaro perchè la funzione Q^2 introdotta nel paragrafo 3.4.2 ha funzione di distribuzione di di χ^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 è

$$Z = (\vec{X} - \vec{\mu})^T V^{-1} (\vec{X} - \vec{\mu}) \tag{76}$$

ad avere funzione di distribuzione di di χ^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 condsideriamo 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, ad esempio, la distribuzione di Maxwell-Bolzman delle velocità molecolari di un gas, e da questa la velocità media, ecc (vedi appunti).

4.4 Somma di variabili casuali

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

* dimostrazioni

———-

4.5 Media aritmetica

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

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{77}$$

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

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

$$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \mu \tag{78}$$

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

$$var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n}, \tag{79}$$

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 μ, cioè che

$$\lim_{n\to\infty} P(|\bar{X} - \mu| \geq \epsilon) = 0 \tag{80}$$

dove ϵ è un numero positivo piccolo a piacere.

* dimostrazione per σ^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 $n \to \infty$, $\bar{X}$ ha distribuzione $N(\mu, \sigma^2/n)$ qualunque sia $f(X)$ purchè σ^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 intervangono 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

$$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} \mu_i \;\text{ e }\; var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2. \tag{81}$$

4.6 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

$$Z = \sum_{i=1}^{n} (X_i - \mu)^2 \tag{82}$$

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

$$E[Z] = \sigma^2 E[\chi^2_n] = n\sigma^2\,, \quad \sigma^2_Z = (\sigma^2)^2 var(\chi^2_n) = 2n(\sigma^2)^2. \tag{83}$$

Nel caso in cui non si conosca μ, si può introdurre la varabile

$$Z = \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{84}$$

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

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{n}{n-1} (\overline{X^2} - \overline{X}^2) \tag{85}$$

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 χ^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. (85) permettono di costruire stime centrate della covarianza di due variabili casuali. Date due variabili X, Y con valori di aspettazione μ_x, μ_y, la loro covarianza è

$$cov(X,Y) = E[(X - \mu_x)(Y - \mu_y)]\,. \tag{86}$$

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

$$v_{x,y} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n-1} (\overline{XY} - \overline{X} \cdot \overline{Y}) \tag{87}$$

ha come valore di apettazione $cov(X,Y)$, ed è quindi uno stimatore centrato della covarianza.

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

$$r = \frac{v_{x,y}}{\sqrt{s_X^2 s_Y^2}} = \frac{\overline{XY} - \overline{X}\cdot\overline{Y}}{\sqrt{(\overline{X^2} - \overline{X}^2)(\overline{Y^2} - \overline{Y}^2)}}. \tag{88}$$

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

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

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 σ_{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 $k = 1, 2, 3$.

5 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 frequenzista e l'inferenza bayesana, che si differenziano essenzialmente per l'interpretazione di probabilità e di conseguenza per l'utilizzo e l'importanza che si d al teorema di Bayes.

Nella statistica frequenzista, 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 statistci, 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 risutati possono ovviamente essere molto diversi.

6 Intervalli di confidenza

Se lo scopo di un esperimento è determinare un parametro θ 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 θ e che venga utilizzata come stimatore la statistica t, funzione del campione, il cui valore di aspettazione sia il valore vero θ_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 γ, per ogni valore di θ si può determinare un intervallo $[t_a, t_b]$ tale che

$$\int_{t_a}^{t_b} f(t)dt = \gamma \tag{89}$$

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

$$\gamma = P(t_a^* < t < t_b^*) = P(\theta_a^* < \theta < \theta_b^*) \tag{90}$$

il corrispondente intervallo per θ, $[\theta_a^*, \theta_b^*]$, è detto intervallo di confidenza con "livello di confidenza" γ. 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 γ 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. (89) 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 probabiltà 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 distribuzione 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 7).

Il collegamento con il test d'ipotesi (capitolo 8) sarà chiaro nel seguito.

6.1 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 μ sia il valore della grandezza fisica che si vuole determinare e la varianza σ^2) sia dovuta alla precisione del metodo di misura. Avendo a disposizione un campione costituito da n misure x_i indipendenti, lo stimatore $\bar{x} = (\Sigma_i x_i)/n$ 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

$$y = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{91}$$

ha distribuzione normale standard e la probabilità che il valore misurato di y cada in $(-1, 1)$ (cioè che $\bar{x}$ differisca da μ 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, di 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 χ^2_1.

Varianza non nota

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

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 \tag{92}$$

possiamo introdurre, oltre a y definita in eq.(91), la variabile causale

$$z = \frac{(n-1)s^2}{\sigma^2} \tag{93}$$

che è una variabile casuale di χ^2_{n-1}. Il rapporto tra una variabile casuale con distribuzione $N(0,1)$ e la radice quadrata di una variabile casuale con distribuzione χ^2_ν 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 ν gradi di libertà:

$$f(t,\nu) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu}\Gamma[\nu/2]} \left(1 + \frac{t^2}{\nu^2}\right)^{-(\nu+1)/2} . \tag{94}$$

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 $\nu \to \infty$ è la funzione di distribuzione di Gauss.

Con riferimento alle eq. (91) e (93), possiamo quindi introdurre la variabile casuale

$$t = \frac{y}{\sqrt{z/(n-1)}} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{95}$$

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 cel caso precedente per detrminare gli intervalli di confidenza.

Da notare che nella definizione di t non compare più esplicitamente σ, che quindi non deve essere detarminato per ottenere gli intervalli di confidenza per μ. 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.

7 Stima dei parametri

7.1 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 decrive 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 capione. 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: $\hat{\theta} = t(X_1, X_2, ...X_n)$, dove $X_1, X_2, ...X_n$ 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: $\hat{\theta}_0 = t(x_1, x_2, ...x_n)$. 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

$$f(X_1, X_2, ...X_n, \vec{\hat{\theta}}) = \prod_{i=1}^{n} f(X_i, \vec{\hat{\theta}}). \tag{96}$$

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 $x_1, x_2, ... x_n$ è una funzione dei soli parametri e può essere pensata come la funzione densità di probabilità dei parametri.

7.2 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 decritte alcune delle caratteristiche probabilistiche che un buon stimatore dovrebbe possedere. Verrà considerato il caso di un parametro, e lo stimatore, funzione del campione $X_1, X_2, ...X_n$, verrà indicato con t.

Consistente
Lo stimatore t è "consistente" se converge in probabilità al valore vero del parametro. Ciò significa che, se $\hat{\theta}_n = t(x_1, x_2, ...x_n)$ è il valore che lo stimatore assume su uno specifico campione e θ_0 il valore vero parametro, è

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta_0| \leq \epsilon) = 1. \tag{97}$$

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

$$P(|\hat{\theta}_n - \theta_0| \leq \epsilon) \geq \eta, \tag{98}$$

con ϵ e η numeri positi 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 unbiassed, 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

$$\lim_{n\to\infty} b = 0, \tag{99}$$

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 mediano, ed è per questo che in genere viane 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 $\tau(\theta)$ é una funzione del parametro θ (con eventualmente $\tau(\theta) = \theta$) di cui la statistica $t(\vec{X})$ è uno stimatore, vale la diseguaglianza di Cramer-Rao (o di Cramer-Rao-Frechèt):

$$\sigma_t^2 \quad \geq \quad \frac{\left(\frac{d\tau}{d\theta} + \frac{db}{d\theta}\right)^2}{E\left[-\frac{\partial^2 lnL}{\partial\theta^2}\right]} \tag{100}$$

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 θ 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:

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

ed è anche

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

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

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

moltiplicando per τ si ottiene

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

Sottraendo le due espressioni trovate si ottiene

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

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

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

e quindi

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

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

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

da cui eq. (100).

———-

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

Notare che

- il lnL deve essere derivabile due volte rispetto a θ;
- se $b = 0$ eq. (100) ha un'espressione più semplice e se, come spesso accade, è anche $\tau(\theta) = \theta$, la diseguaglianza diventa

$$\sigma_t^2 \quad \geq \quad \frac{1}{E\left[-\frac{\partial^2 lnL}{\partial\theta^2}\right]} \tag{101}$$

- l'uguaglianza vale se e solo se

$$\frac{\partial lnL}{\partial\theta} = k\,(t-\tau) \tag{102}$$

  dove k può anche essere funzione di θ.

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 $\epsilon_t = \sigma^2_{min}/\sigma^2_t$, 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'e' la "sufficienza": uno stmatore 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.

7.3 Metodo del Maximum Likelihood

Questo metodo (MML), molto generale e diffuso, permette di determinare stimatori con proprietà richieste. In condizioni molto genrali 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.

7.3.1 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 θ è il parametro da stimare a partire da un insieme di misure indipendenti $X_1, X_2, ...X_n$. Per un certo θ la funzione densità di probabilità congiunta per le variabili X_i è

$$L(\vec{X}, \theta) = \prod_{i=1}^{n} f(X_i, \theta). \tag{103}$$

che può anche essere vista cone la funzione densità di probabilità del parametro θ 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 θ 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:

$$\frac{\partial L}{\partial \theta} = 0 \quad \text{oppure} \quad \frac{\partial lnL}{\partial \theta} = 0 \tag{104}$$

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 $x_1, x_2, ...x_n$, è la stima del parametro $\hat{\theta} = t(\vec{x})$. Naturalmente bisogna verificare che sia

$$\left[\frac{\partial^2 lnL}{\partial \theta^2}\right]_{t=E[t]} < 0\,. \tag{105}$$

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

$$\frac{\partial lnL}{\partial \theta_j} = 0\,. \tag{106}$$

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 lnL numericamente (come fatto dai programmi di fit più diffusi, come Minuit).

7.3.2 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

$$f(X,\tau) = \frac{1}{\tau} e^{-X/\tau} \tag{107}$$

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

$$L(\vec{X},\tau) = \frac{1}{\tau} \prod_{i=1}^{n} e^{-X_i/\tau} \,. \tag{108}$$

Ponendo la derivata di lnL rispetto a τ uguale a 0 si ottiene che lo stimaore di τ è

$$t_\tau = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{109}$$

ed è

$$E\left[\frac{\partial^2 L}{\partial \theta^2}\right]_{t=E[t]} = E\left[\frac{n}{\tau^3}(\tau - 2n\bar{X})\right]_{t=E[t]} = -\frac{n}{\tau^2} < 0 \,, \tag{110}$$

essendo $E[X_i] = \tau$.
Se si fosse ipotizzata la funzione di distribuzione

$$f(X,\lambda) = \lambda e^{-\lambda X} \tag{111}$$

seguendo lo stesso procedimento avremmo ottenuto come stimatore di λ 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_θ e t_a si ottengono rispettivamente da

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

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 di τ e λ si ha

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

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_τ è a varianza minima.

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

- stima di μ nel caso di stesso valore di aspettazione e stessa varianza nota ($N(\mu, \sigma^2)$)
- stima di μ nel caso di stesso valore di aspettazione ma varianze diverse e note ($N(\mu, \sigma_i^2)$)
- stima di σ^2 nel caso di stesso valore di aspettazione noto e stessa varianza ($N(\mu, \sigma^2)$)
- stima simultanea di μ e σ^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 τ nel caso in cui gli intervalli di tempo misurati siano limitati all'intervallo $(0, T)$.

2.Dimostrare che:
- le stime di μ e σ^2 corrispondono effettivamente a un massimo di L;
- la stima di σ è la radice quadrata della stima di σ^2;
- la stima di σ non è centrata ma è consistente.

7.3.3 Ulteriori considerazioni

Proprietà delle stime

Il MML è molto usato per le propietà di cui godono la 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 gneralmente per la varianza della stima di un parametro θ si usa l'espressione

$$\sigma_{\hat{\theta}}^2 = \frac{1}{\left[-\frac{\partial^2 lnL}{\partial \theta^2}\right]_{\theta=\hat{\theta}}} \tag{112}$$

- 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 (106). Gli elementi della matrice V delle covarianze si ottengono da

$$(V^{-1})_{ij} = \left[-\frac{\partial^2 lnL}{\partial\theta_j \partial\theta_k} \right]_{\vec{\theta}=\vec{\hat{\theta}}}. \tag{113}$$

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:

$$L(t) = L_0 \, e^{-\frac{(t-\theta_0)^2}{2\sigma_t^2}} \tag{114}$$

dove t è lo stimatore del parametro e θ_0 il suo valore. Data la stima $\hat{\theta}$, i possibili valori delle stime hanno quindi distribuzione

$$L(\theta) = L_{max} e^{-\frac{(\theta-\hat{\theta})^2}{2\sigma_{\hat{\theta}}^2}} \tag{115}$$

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

$$lnL(\theta) = lnL_{max} - \frac{(\theta-\hat{\theta})^2}{2\sigma_{\hat{\theta}}^2} \tag{116}$$

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

$$\theta - \hat{\theta} = \pm\sigma_{\hat{\theta}} \, . \tag{117}$$

Per ottenere la stima del parametro e la sua deviazione standard si può quindi procedere graficamente (o numericamente): si calcola lnL sul campione per diversi valori di θ, si riportanto in un grafico i valori ottenuti e si determina la parabola. Il valore massimo individua la stima $\hat{\theta}$ e i valori di θ corrispondenti a $lnL_{max} - 1/2$ permettono di determinare $\sigma_{\hat{\theta}}$, cioè l'intervallo corrispondente a un livello di confidenza di 0.68 (distribuzione di Gauss), i valori di θ corrispondenti a $lnL_{max} - 2$ 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. (116) si può ottenere anche sviluppando in serie lnL in un intorno della stima:

$$lnL(\theta) = lnL(\hat{\theta}) + \left[\frac{\partial lnL}{\partial\theta} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2} \left[\frac{\partial^2 lnL}{\partial\theta^2} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + ... \tag{118}$$

Il primo termine è lnL_{max}, il secondo è nullo perchè la stima corrisponde al massimo di lnL ed il terzo è il secondo di eq. (116), nell'ipotesi si stime efficienti.

Nel caso di k parametri, generalizzando quanto visto al caso di variabili con funzione di distribuzione multinormale, eq. (116) diventa

$$lnL(\vec{\theta}) = lnL_{max} - \frac{1}{2}(\vec{\theta} - \vec{\hat{\theta}})^T V^{-1} (\vec{\theta} - \vec{\hat{\theta}}) \tag{119}$$

Il valore lnL_{max} corrisponde alle stime $\vec{\hat{\theta}}$. I valori di $lnL(\vec{\theta}) = lnL_{max} - 1/2$ 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 lnL). Per $k = 2$, si fissano diversi valori ad esempio di θ_1 e si determina per ciascuno la parabola in funzione di θ_2. I massimi di lnL, riportati in funzione di θ_1 si trovano ancora su una parabola il cui massimo è lnL_{max}. I punti corrispondenti a $lnL_{max} - 1/2$ 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 paramteri 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

7.4 Metodo dei Minimi Quadrati

7.4.1 Principio

Anche questo metodo (MMQ o "least squared method", LSM) permette di ottenere stime con molte delle proprietà auspicabili elencate in precdenza. 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 $(z_1, z_2, ...z_N)$ avremo N valori misurati di Y $(y_1, y_2, ...y_N)$ 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

$$X^2 = \sum_{i=1}^{N} w_i (y_i - \mu_i)^2 \tag{120}$$

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 $\hat{\theta}_1, \hat{\theta}_2, ...\hat{\theta}_k$ 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 $T = c\, l^{1/2}$ 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

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

dove σ_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 vogliano determinare i paramtri 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, μ_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 σ_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

$$X^2 = (\vec{y} - \vec{\mu})^T V^{-1} (\vec{y} - \vec{\mu}) \tag{121}$$

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 μ_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, mantendendo il ruolo iniziale di $\vec{z}$ e $\vec{y}$, si possono sostituire le $\sigma^2_{y_i}$ con $\sigma^2_i = \sigma^2_{y_i} + \sigma^2_{\mu_i}$ dove $\sigma^2_{\mu_i}$ è la varianza di μ_i ottenita da $\sigma^2_{z_i}$ con la legge di propagazione della varianza. Le varianze $\sigma^2_{\mu_i}$ 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^2_{min} è una variabile casuale con funzione di distribuzione di χ^2 ed è per questo motivo che il metodo viene anche chiamato minimizzazione del χ^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.

7.4.2 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:

$$\mu_i = f(z_i, \vec{\theta}\,) = \sum_{j=1}^{k} a_{ij}(z_i)\,\theta_j\,, \quad \text{o} \quad \vec{\mu} = A\vec{\theta}\,, \tag{122}$$

e i pesi (le varianze $\sigma^2_{y_i}$) 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 σ^2_i per y_i. In questo caso la funzione dei parametri da minimizzare è

$$X^2 = \sum_{i=1}^{n} \frac{[y_i - (mz_i + q)]^2}{\sigma^2_i}. \tag{123}$$

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

$$\hat{m} = \frac{S_{00}S_{11} - S_{10}S_{01}}{S_{00}S_{20} - S^2_{10}}\,, \quad \hat{q} = \frac{S_{20}S_{01} - S_{10}S_{11}}{S_{00}S_{20} - S^2_{10}} \tag{124}$$

dove

$$S_{00} = \sum_{i=1}^{n} \frac{1}{\sigma^2_i}\,,\ \ S_{01} = \sum_{i=1}^{n} \frac{y_i}{\sigma^2_i}\,,\ \ S_{10} = \sum_{i=1}^{n} \frac{z_i}{\sigma^2_i}\,,\ \ S_{11} = \sum_{i=1}^{n} \frac{z_i y_i}{\sigma^2_i}\,,\ \ S_{20} = \sum_{i=1}^{n} \frac{z^2_i}{\sigma^2_i}\,. \tag{125}$$

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

$$\sigma^2_{\hat{m}} = \frac{S_{00}}{S_{00}S_{20} - S_{10}^2}\,, \quad \sigma^2_{\hat{q}} = \frac{S_{20}}{S_{00}S_{20} - S_{10}^2}\,, \quad cov(\hat{m}, \hat{q}) = -\frac{S_{10}}{S_{00}S_{20} - S_{10}^2}\,. \tag{126}$$

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 $Y = a_0 + a_1 Z + a_2 Z^2$ si ha

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

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

La forma quadratica da minimizzare è data in eq. (121) e può essere riscritta come

$$X^2 = (\vec{y} - A\vec{\theta})^T V^{-1} (\vec{y} - A\vec{\theta}). \tag{127}$$

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

$$\vec{\hat{\theta}} = B\,\vec{y} \quad \text{con} \quad B = (A^T V^{-1} A)^{-1} A^T V^{-1} \tag{128}$$

* infatti:

$$\frac{\partial X^2}{\partial \vec{\theta}} = -2(\vec{y} - A\vec{\theta})^T V^{-1} A\,. \tag{129}$$

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.(128).
———-

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 $V_{\vec{\hat{\theta}}} = (A^T V^{-1} A)^{-1}$ e B può anche essere scritta nella forma

$$B = V_{\vec{\hat{\theta}}} A^T V^{-1}. \tag{130}$$

* 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 $Y = a_0 + a_1 Z + a_2 Z^2$, 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.

7.4.3 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. (127) e possono essere scritte come

$$\vec{\hat{\theta}} = V_{\vec{\hat{\theta}}} A^T V^{-1} \vec{y}\,. \tag{131}$$

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

$$\begin{aligned} X^2 &= (\vec{y} - A\vec{\hat{\theta}})^T V^{-1} (\vec{y} - A\vec{\hat{\theta}}) + (\vec{\hat{\theta}} - \vec{\theta})^T A^T V^{-1} A (\vec{\hat{\theta}} - \vec{\theta}) + \\ &\quad + (\vec{y} - A\vec{\hat{\theta}})^T V^{-1} A (\vec{\hat{\theta}} - \vec{\theta}) + (\vec{\hat{\theta}} - \vec{\theta})^T A^T V^{-1} (\vec{y} - A\vec{\hat{\theta}})\,. \end{aligned} \tag{132}$$

Il primo termine è il valore minimo di X^2

$$X^2_{min} = (\vec{y} - A\vec{\hat{\theta}})^T V^{-1} (\vec{y} - A\vec{\hat{\theta}})\,), \tag{133}$$

il secondo è $(\vec{\hat{\theta}} - \vec{\theta})^T V_\theta^{-1} (\vec{\hat{\theta}} - \vec{\theta})$, 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

$$X^2 = X^2_{min} + (\vec{\hat{\theta}} - \vec{\theta})^T V_\theta^{-1} (\vec{\hat{\theta}} - \vec{\theta})\,, \tag{134}$$

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

di χ^2_ν 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. (134) è un paraboloide in genere deformato, così come gli elissoidi 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^2_{min}, sia le regioni di confidenza, date da $X^2 = X^2_{min} + 1, +4, ...$

8 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 statististico ("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_o se è vera un'ipotesi alternativa H_1, e minimizza la probabilità di rigettare l'ipotesi nulla se corretta.

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

- Si introduce una opportuna statistica, cioè una funzione del campione, $t(x_1, ...x_n)$ che è detta "statistica di test" e della quale nell'ipotesi H_o (e H_1) si conosce la funzione di distribuzione $\Phi_o(t)$. Ad esempio, la funzione di distribuzione di t nel caso in cui sia vera l'ipotesi nulla H_o potrebbe essere la funzione $\Phi_o(t)$ riportata in fig. 4. Se invece fosse vera un'ipotesi alternativa H_1 potrebbe essere la funzione $\Phi_1(t)$.

- Si sceglie un "livello di significatività" α per il test d'ipotesi, in genere 0.10, o 0.05, o 0.01. Essendo $\Phi_o(t)$ nota, dal valore di α si calcola t_α tale che

$$\int_{t_\alpha}^{t_{max}} \Phi_o(t)dt = \alpha. \tag{135}$$

  cioè tale che la probabiltà che t sia maggiore di t_α sia α. L'intervallo (t_α, t_{max}) è detto "regione critica".

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

Minore è α, minore è la probabilità di rigettare H_o se corretta (errore del I tipo). Al diminuire di α aumenta però la probabilità β

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

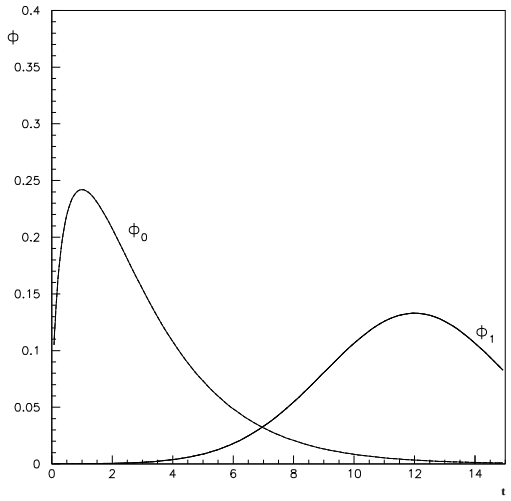

Figure 4: Funzioni di distribuzione $\Phi(t)$ vs t per due diverse ipotesi.

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 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 α, minimizza β. 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 β 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 α^* con $\alpha^* = \int_{t^*}^{t_{max}} \Phi_o(t)dt$;
- è stato considerato il test "a una coda", avendo introdotto la regione critica (t_α, t_{max}). Il test a "due code" è trattato negli esempi descritti nel seguito.

8.1 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 $x_1, ...x_n$ 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 μ e σ^2 e le n osservazioni disponibili. Gli esempi che seguono sono per certi aspetti simili a quanto visto nel paragrafo 6.1.

Test della media Nel caso in cui la varianza σ^2 sia nota, si vuole verificare se il campione è in accordo con l'ipotesi nulla H_o: $\mu = \mu_0$ con μ_0 valore previsto del valore di aspettazione. Come ipotei 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

$$t(x_1, ...x_n) = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}. \tag{136}$$

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 α, 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 significtività pari a $\alpha/2$. Si definisce quindi $t_{\alpha/2} > 0$ tale che

$$\int_{-\infty}^{-t_{\alpha/2}} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} e^{-t^2/2} dt = \int_{t_{\alpha/2}}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}} e^{-t^2/2} dt = \frac{\alpha}{2} \tag{137}$$

e si rigetta H_o con livello di significatività α 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 α, $t_{\alpha/2}$ 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 6.1

Test della varianza

Supponiamo che sul valore noto di μ non ci siano dubbi e che si voglia verificare l'accordo del campione con il valore previsto della varianza σ_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

$$t(x_1, ...x_n) = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma_0^2} \tag{138}$$

ha funzione di distribuzione di χ^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 α, dai valori tabulati della funzione cumulativa di χ^2, si trovano i valori $t^-_{\alpha/2}$ e $t^+_{\alpha/2}$ tali che gli integrali della funzione di distribuzione di χ^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 χ^2 è molto asimmetrica per n piccolo.

L'ipotesi H_o verrà rigettata se t^*, il valore della statisica 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.

8.2 Test di χ^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 σ_i^2 note. Supponiamo di voler verificare che i dati siano in accordo con certi valori di aspettazione μ_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

$$t(y_1, ...y_n) = \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2} \tag{139}$$

ha funzione di distribuzione di χ^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 α il valore critico t_α tale che

$$\int_{t_\alpha}^{\infty} f(t)dt = \alpha \tag{140}$$

dove $f(t)$ è la funzione di distribuzione di χ^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. (139) non ha funzione di distribuzione di χ^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

   $$t(y_1, ...y_n) = (\vec{y} - \vec{\mu})^T V^{-1} (\vec{y} - \vec{\mu}) \tag{141}$$

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

3. Se i valori μ_i sono stati ottenuti utilizzando lo stesso campione usato per il test d'ipotesi, t ha funzione di distribuzione di χ^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 μ_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 del minimi quadrati ha la stessa espressione della statistica t e, nelle stesse condizioni, è una variabile casuale con funzione di distribuzione di χ^2 e il valore di X^2_{min} ottenuto permette di eseguire direttamente anche il test d'ipotesi.

4. Se ν è sufficientemente grande la funzione di distribuzione di χ^2 tende a funzione di distribuzione di Gauss con valore di aspettazione ν e varianza 2ν. Questo permette di avere immediatamente un'idea dell'accordo tra campione e modello, senza calcolare t_α. Inoltre, in media ci si aspetta che ciascun addendo contribuisca al χ^2 con circa 1: se un contributo è maggiore di 9 (differenza maggiore di 3 σ_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 probabiltà

$$p^*_j = \int_0^{t^*_j} f_j(t)dt$$

dove $f_j(t)$ è la funzione di distribuzione di χ^2 con ν_j gradi di libertà, con la distribuzione uniforme.

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

Il test di χ^2 ha delle limitazioni tra le quali il fatto di essere poco selettivo se ν 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 μ è noto, la statistica

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

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

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

con $\bar{y}$ media pesata (o aritmetica, se i valori σ_i^2 sono uguali tra loro), ha ancora funzione di distribuzione di χ^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 σ_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 $Y = f(X; a_1, ..., a_K)$, con $a_1, ...a_k$ parametri noti. In questo caso l'ipotesi nulla è H_o: $\mu_i = E[y_i] = f(x_i; a_1, ..., a_K)$, $i = 1, n$ e, se è vera, la statistica t di eq. (139) ha funzione di distribuzione di χ^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 ν, si fissano α e t_α, si calcola t^* e, in caso, si rigetta l'ipotesi che tra X e Y ci sia la relazione $Y = f(X; a_1, ..., a_K)$.

Altro caso importante di test di χ^2 è quello in cui si ha a disposizione un campione costituito da n valori x_i di una variable 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 Δ_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 Δ_k è sufficientemente piccolo. Se i valori μ_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

$$t(n_1, ...n_N) = (\vec{n} - \vec{\mu})^T V^{-1} (\vec{n} - \vec{\mu}) \,, \tag{142}$$

o

$$t(n_1, ...n_N) = \sum_{k=1}^{N} \frac{(n_k - \mu_k)^2}{\mu_k} \tag{143}$$

se le covarianze sono trascurabili, ha funzione di distribuzione di χ^2. Il numero di gradi di libertà è $\nu = N - 1$ se i parametri che compaiono nella funzione $f(x)$ sono noti: c'e' 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 χ^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.

8.3 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 costrisce cosi 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 χ^2.

* per una descrizione più completa ed esempi vedi appunti

Contents

1 Momenti **3**
- 1.1 Momenti algebrici e centrali . . . 3
- 1.2 Esempi di funzioni generatrici dei momenti . . . 4
  - 1.2.1 Distribuzione binomiale . . . 4
  - 1.2.2 Funzione di distribuzione di Gauss . . . 5

2 Alcune distribuzioni e densità di probabilità **6**
- 2.1 Distribuzione di Poisson . . . 6
- 2.2 Distribuzione dei tempi di attesa . . . 8
- 2.3 Distribuzione Gamma . . . 10
- 2.4 Distribuzione di χ^2 . . . 10
- 2.5 Distribuzione di Cauchy . . . 11

3 Più variabili casuali **13**
- 3.1 Covarianza . . . 13
- 3.2 Funzioni generatrici dei momenti . . . 14
- 3.3 Esempi di variabili casuali correlate . . . 15
- 3.4 Alcune distribuzioni congiunte . . . 16
  - 3.4.1 Distribuzione multinomiale . . . 16
  - 3.4.2 Distribuzione multinormale . . . 18

4 Funzioni di variabili casuali **22**
- 4.1 Funzioni di una o più variabili casuali . . . 22
- 4.2 Popolazione, campione e funzioni di variabili casuali . . . 23
- 4.3 Somma di quadrati di variabili casuali . . . 24
- 4.4 Somma di variabili casuali . . . 25
- 4.5 Media aritmetica . . . 25
- 4.6 Somme di quadrati degli scarti . . . 26

5 Inferenza statistica **29**

6 Intervalli di confidenza **30**
- 6.1 Misure con distribuzione di Gauss . . . 31

7 Stima dei parametri **33**
- 7.1 Considerazioni generali . . . 33
- 7.2 Caratteristiche degli stimatori . . . 34
- 7.3 Metodo del Maximum Likelihood . . . 37
  - 7.3.1 Stimatori . . . 37
  - 7.3.2 Esempi . . . 38
  - 7.3.3 Ulteriori considerazioni . . . 39
- 7.4 Metodo dei Minimi Quadrati . . . 41
  - 7.4.1 Principio . . . 41

7.4.2 Il metodo dei Minimi Quadrati Lineare 43
7.4.3 Incertezze statistiche 45

8 Test d'ipotesi **47**
8.1 Test parametrici per variabili gaussiane 48
8.2 Test di χ^2 50
8.3 Test di indipendenza 52