

Appunti Laboratorio

Pietro Faraggiana

July 2020

Capitolo 1

Misure ed incertezze

Analisi dimensionale

Grazie all'analisi dimensionale è possibile controllare od addirittura intuire le equazioni cercate. Esempio con il periodo del pendolo:

le quantità che possiamo immaginare influiscano sul tempo sono massa, gravità e lunghezza del filo, quindi:

$$T = l^a m^b g^c = l^a m^b t^{-2c} l^c = l^{a+c} m^b t^{-2c}$$

Dato che il periodo ha come unità di misura la sola unità temporale allora l'equazione deve avere t^1 come unità da ambo i lati, si arriva quindi ad un sistema del genere:

$$\begin{cases} a + c = 0 \\ b = 0 \\ -2c = 1 \end{cases}$$

Risolvendo il sistema si arriva alla corretta formula per il periodo del pendolo.

Errori di misura

$$x = \tilde{x} \pm \Delta x [Unità di misura]$$

Questa è la formula generale, quello più utilizzato è l'errore satistico ma esistono anche errore massimo ed errore relativo. 2 errori sono detti compatibili se l'intersezinone tra i due insiemi dati dalle zone di incertezza è non nulla.

Errore massimo

$$x = \tilde{x} \pm \Delta x [Unità di misura]$$

- Δx è l'incertezza di misura, in questo caso *l'errore massimo* ovvero il più piccolo intervallo in cui la misura può ricadere (con al massimo 2 cifre significative), l'errore massimo è spesso inutilizzato poichè nuove misurazioni possono confermarlo od allargarlo.
- x è l'oggetto della nostra ricerca.
- \tilde{x} è invece la stima centrale oppure la miglior stima.

Errore relativo

$$\text{Errore relativo} = \frac{\Delta x}{|\tilde{x}|}$$

L'errore relativo è il rapporto tra l'incertezza di misura e la miglior stima.

Errore statistico approssimato

Fino alle probabilità questa formula sarà considerata come errore statistico approssimativo:

$$\sigma_x = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \tilde{x})^2}$$

Precisione ed accuratezza

La precisione è l'ampiezza dell'incertezza di misura, più essa è larga meno la misura è precisa; l'accuratezza è la vicinanza delle misurazioni alla vera misura.

Cifre significative

Le cifre significative sono definite da 3 regole:

- La cifra più significativa è quella più a sinistra diversa da zero.
- La cifra meno significativa è quella più a destra (inclusi gli zeri a destra del separatore decimale, ma fatta eccezione per gli zeri che definiscono l'ordine di grandezza per i numeri interi).
- Tutte le cifre comprese tra la più significativa e la meno significativa sono significative.

Propagazione dell'errore

Propagazione dell'errore massimo nel caso della somma e della differenza

$$S = (\tilde{x} + \tilde{y}) \pm (\Delta x + \Delta y)$$
$$D = (\tilde{x} - \tilde{y}) \pm (\Delta x + \Delta y)$$

Propagazione dell'errore massimo nel caso del prodotto e del rapporto

$$\tilde{P} = \tilde{x}\tilde{y} + \Delta x\Delta y$$
$$\Delta P = |\tilde{y}|\Delta x + |\tilde{x}|\Delta y$$

Molto spesso $\Delta x\Delta y$ è trascurabile data la sua piccolezza

$$\tilde{Q} \approx \frac{\tilde{x}}{\tilde{y}}$$
$$\Delta Q \approx \frac{|\tilde{x}|\Delta y + |\tilde{y}|\Delta x}{y^2}$$

Nel caso si trattasse invece di errori relativi nel prodotto e nel rapporto essi si sommano.

Propagazione dell'errore massimo nel caso generale unidimensionale

I casi appena fatti sono la propagazione nel caso delle funzioni elementari, nel caso di altre operazioni (o meglio, funzioni) possiamo trasformare la funzione in un polinomio tramite gli sviluppi di Taylor:

$$f(\tilde{x} \pm \Delta x) \approx f(\tilde{x}) \pm \frac{df}{dx}(\tilde{x})\Delta x$$
$$\tilde{f} \approx f(\tilde{x})$$
$$\Delta f \approx \left| \frac{df}{dx}(\tilde{x}) \right| \Delta x$$

Propagazione dell'errore massimo nel caso generale

Nel caso generale l'unica cosa che cambia dal caso generale unidimensionale è che si possono appunto avere più dimensioni, quindi al posto della derivata semplice usiamo la derivata parziale:

$$\Delta x \approx \left| \frac{\partial f}{\partial x}(\tilde{x}, \tilde{y}, \tilde{z} \dots) \right| \Delta x + \left| \frac{\partial f}{\partial y}(\tilde{x}, \tilde{y}, \tilde{z} \dots) \right| \Delta y + \left| \frac{\partial f}{\partial z}(\tilde{x}, \tilde{y}, \tilde{z} \dots) \right| \Delta z \dots$$

Ricetta per la propagazione dell'errore statistico

Quanto detto fin'ora è inutile ai fini pratici perché l'errore massimo è inutilizzato. La propagazione dell'errore statistico è

$$\sigma_f^2 \approx \left(\frac{\partial f}{\partial x}(\tilde{x}, \tilde{y}, \tilde{z} \dots) \right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}(\tilde{x}, \tilde{y}, \tilde{z} \dots) \right)^2 \sigma_y^2 + \left(\frac{\partial f}{\partial z}(\tilde{x}, \tilde{y}, \tilde{z} \dots) \right)^2 \sigma_z^2 \dots$$

Capitolo 2

rappresentazione dati

In Fisica si ha spesso a che fare con enormi quantità di dati che non sono facilmente interpretabili nella loro forma grezza. Rappresentare opportunamente tali dati ed estrarre l'informazione rilevante in una forma intelligibile è uno degli compiti fondamentali del Fisico sperimentale. In questo capitolo illustriamo brevemente alcune idee fondamentali sull'argomento che svilupperemo ulteriormente nel seguito. Ecco alcune forme di raccoglimento dati e formattazione:

- Tabelle
- Grafici di dispersione o grafici xy
- Fit (miglior approssimazione grafica dei dati su di un grafico di dispersione)
- Grafici a barre
- Istogrammi

Scale non lineari

È molto facile per l'occhio umano notare su di un piano cartesiano un discostamento da una funzione lineare, non è altrettanto facile quando la funzione è ad esempio x^2 o x^4 . Per questo motivo si usano altre scale cambiando le variabili.

scala logaritmica

Una scala logaritmica è una scala dove l'unità di misura è la decade. Una decade è un intervallo di valori numerici i cui estremi stiano tra loro nel rapporto di 1 a 10. Dati allora due numeri x_1 ed x_2 , la loro distanza in decadi è data da

$$dex(x_1, x_2) = \log\left(\frac{x_2}{x_1}\right)$$

quindi ad esempio la distanza tra 1 e 2 è $\log(2/1)=0.301$. Ovviamente non è possibile rappresentare lo 0 in scala logaritmica poiché diverge a meno infinito.

Legge di potenza e scala bilogaritmica

Una legge di potenza è una qualsiasi legge del tipo

$$y(x) = Cx^\Gamma$$

Essendo $\log(y) = \Gamma \log(x) + \log(C)$ e ponendo $x' = \log(x)$, $y' = \log(y)$ e $C' = \log(C)$ si ottiene

$$y'(x') = C' + \Gamma x'$$

questa è una retta e come abbiamo detto inizialmente è notevolmente più facile osservare l'andamento di una retta rispetto a quello di una curva. Un piano con scale logaritmiche è detto un in scala bilogaritmica, se un asse è in scala logaritmica e l'altro no allora è detto semi logaritmico. Il coefficiente angolare di tale retta è $\frac{dex(y_1, y_2)}{dex(x_1, x_2)} = \frac{\log y_2 / y_1}{\log x_2 / x_1}$ dove dex indica la decade.

Capitolo 3

Probabilità

Qual è la differenza tra probabilità e statistica? Per inquadrarlo vediamo 2 domande tipiche:

- Domanda 1: Abbiamo un'urna contenente 5 palline, di cui 3 rosse e 2 blu. Se estraiamo (bendati) una pallina qual è la probabilità che essa sia blu?
- Domanda 2: Abbiamo un'urna contenente 5 palline, alcune rosse ed alcune blu (ma non conosciamo la proporzione tra i colori). Estraiamo una pallina, che risulta essere blu, e la mettiamo da parte. Estraiamo una seconda pallina, che stavolta è rossa. Qual è il numero di palline rosse nell'urna?

La prima domanda è facile e si tratta di probabilità, la risposta sarà facilmente $2/5$; la seconda domanda è invece più difficile e si tratta in un certo senso del ragionamento inverso rispetto al primo ed è la statistica (risolvibile con la verosimiglianza), non ci è dato sapere con certezza il risultato ma un gap di risultati con ognuno una probabilità di apparenza.

Definizione di probabilità

La struttura di base su cui si fonda la definizione assiomatica della probabilità (di Kolmogorov) è data dal cosiddetto spazio campionario Ω , cioè l'insieme (che per semplicità assumeremo numerabile) di tutte le possibili realizzazioni elementari di un dato fenomeno. Sulla base di questo si definisce poi lo spazio degli eventi F come l'insieme di tutti i sottoinsiemi di Ω (che si dice anche l'insieme di potenza di Ω). L'idea di base è che la probabilità è definita proprio sullo spazio degli eventi, cioè si può assegnare una probabilità non solo ad un qualsiasi elemento dello spazio campionario, ma anche ad uno qualsiasi dei suoi sottoinsiemi. Siccome l'unione e l'intersezione di sottoinsiemi dello spazio campionario fanno ancora parte dello spazio degli eventi, avrà senso parlare anche della probabilità dell'unione e dell'intersezione di eventi.

teoria assiomatica

$$\begin{aligned} \text{Probabilità evento} &= P(E) & \text{Probabilità evento complementare} &= P(\bar{E}) = 1 - P(E) \\ \text{Unione insiemi} &= P(E_1 \cup E_2) = P(E_1) + P(E_2) - (P(E_1) \cap P(E_2)) & (\text{formula di grassman}) \\ \text{Probabilità condizionata} &= P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \end{aligned}$$

La probabilità condizionata è la probabilità che si verifichi un evento E_1 nel caso in cui sappiamo già che si è verificato l'evento E_2 .

Due eventi sono detti indipendenti se

$$P(E_1|E_2) = P(E_1) \text{ oppure } P(E_2)$$

Da essa viene il teorema di Bayes anche detto della probabilità a posteriori:

$$P(E_i|E_2) = \frac{P(E|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(E|A_j)P(A_j)}$$

Oltre a queste esistono la definizione combinatoriale, frequentista e soggettiva.

Variabili casuali e funzioni di distribuzione

Una variabile casuale rappresenta la realizzazione numerica di un processo casuale, esistono *variabili casuali discrete*, ovvero variabili che in un determinato intervallo sono numerabili, e *variabili casuali continue* ovvero variabili con un numero illimitato di elementi in un intervallo, ed entrambe possono essere finite od infinite

Varabili casuali discrete

Consideriamo una variabile discreta x che può assumere n valori distinti possibili $x_1 \dots x_n$. Indicando con $P(x_k)$ la probabilità che x assuma il valore x_k , definiamo funzione di distribuzione di x la funzione che associa ad ogni valore x_k della variabile x la sua probabilità $P(x_k)$ (funzione rappresentata in maniera tabellare o come grafico a barre). Secondo gli assiomi di Kolmogorov

$$\sum_k P(x_k) = 1$$

Varabili casuali continue

Su una variabile continua ha poco senso chiedersi la probabilità che accada un singolo evento determinando quindi ci si chiede la probabilità su di un intervallo del dominio, come ad esempio (x_0, dx) tale che $P(x_0 \leq x \leq x_0 + dx)$, a questo punto possiamo definire la *densità di probabilità* (in poche parole sarebbe la probabilità per unità di intervallo sotto forma di rapporto incrementale e quindi una derivata)

$$P(x_0) = \lim_{dx \rightarrow 0} \frac{P(x_0, dx)}{dx}$$

Invece la probabilità che la variabile casuale assuma valore compreso in un intervallo dx centrato in x_0 si scrive

$$p(x_0, dx) = p(x_0)dx$$

e, corrispondentemente, la probabilità che essa assuma un valore compreso nell'intervallo finito $[x_1, x_2]$ è data da:

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

dove se x_1, x_2 sono corrispondentemente $-\infty, +\infty$ allora quel integrale vale 1.

Valore di aspettazione

Sia data una funzione $f(x)$ di una variabile casuale x (continua o discreta). Definiamo il valore di aspettazione come

$$E[f(x)] = \begin{cases} \sum_k f(x_k)P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{+\infty} f(x)P(x)dx & \text{per variabili continue} \end{cases}$$

Il valore di aspettazione trasforma una funzione di x in un numero facendo una media in cui i valori di $f(x)$ sono pesati con il valore della funzione di distribuzione di x , il valore di aspettazione è lineare e se la funzione coincide con una costante allora il valore di aspettazione è la costante stessa, in poche parole è la media.

Tendenza centrale e dispersione intorno alla media

Nelle applicazioni pratiche è spesso utile condensare le varie informazioni di una funzione di distribuzione attorno a dei parametri, i più utilizzati sono il valore medio e quanto la funzione si discosti in media dal valore medio.

Media, mediana e moda

- **Media:** La media è il valore di aspettazione di x e se ci trovassimo in una variabile discreta le cui possibilità di uscita del evento siano uguali indipendentemente dalla x allora la media coincide con la media aritmetica.

$$\mu = E[x] = \begin{cases} \sum_k x_k P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{+\infty} x f(x) P(x) dx & \text{per variabili continue} \end{cases}$$

- **Mediana**($\mu_{1/2}$):Definiamo mediana di una distribuzione quel valore $\mu_{1/2}$ della variabile casuale tale che

$$P(x \leq \mu_{1/2}) = P(x \geq \mu_{1/2})$$

ovvero il punto, pesato con la probabilità equidistante dai due estremi dell' intervallo (la probabilità che preso un punto si trovi a destra o a sinistra della mediana è 0.5), infatti ha poco senso trovare la mediana di una casuale discreta poiché il valore potrebbe non esistere inoltre se in un grafico di distribuzione media e mediana coincidono allora il grafico è simmetrico.

- **Moda**:La moda è semplicemente il valore scelto di più, ovvero, in un grafico di distribuzione è il massimo.

Varianza e deviazione standard

Caratterizzare la dispersione attorno alla media di una variabile casuale significa definire una funzione il cui valore di aspettazione sia una misura di quanto la variabile x tenda a discostarsi dal proprio valor medio. Potremmo essere tentati a porre $f(x)=x-\mu$ ma il risultato sarebbe 0 ($E[x - \mu] = E[x] - \mu = 0$), per ovviare questo problema utilizziamo la funzione $f(x)=(x - \mu)^2$ e definiamo la varianza:

$$\sigma^2 = E[(x - \mu)^2] = \begin{cases} \sum_k (x_k - \mu)^2 P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{+\infty} (x - \mu)^2 P(x) dx & \text{per variabili continue} \end{cases}$$

A questo punto ricordiamoci di farne la radice per trovare la deviazione standard, quindi:

$$\sigma = \sqrt{\sigma^2}$$

Se la varianza è moltiplicata per una costante il risultato è $\text{Var}(cx) = c^2 \text{Var}(x)$ Inoltre dalla definizione di varianza tramite qualche calcolo algebrico si arriva alla conclusione che $\sigma^2 = E[x^2] - \mu^2$. La deviazione standard rappresenta l'entità delle fluttuazioni attorno al valor medio e fornisce una misura conveniente delle fluttuazioni stesse, in poche parole è la media delle distanze dalla misura media di tutte le misure.

La larghezza a metà altezza

Se assumiamo che la distribuzione sia unimodale, cioè che abbia un solo massimo, la retta orizzontale che interseca l'asse delle ordinate in corrispondenza della metà del valore del punto di massimo della distribuzione intersecherà la distribuzione stessa esattamente in due punti x_a ed x_b . Definiamo allora la larghezza a metà altezza FWHM (in inglese) come

$$FWHM = |x_b - x_a|$$

e per trovare la semi-larghezza dividiamo il tutto per 2. In gran parte delle distribuzioni continue di interesse pratico la semilarghezza è $HWHM = c\sigma$ quindi possiamo ora osservare la deviazione standard come *larghezza di una distribuzione*

La disuguaglianza di Chebyshev ed il significato della deviazione standard

Il significato profondo della deviazione standard è sostanzialmente che, se misuriamo le deviazioni dal valor medio in unità di σ , è poco probabile che esse siano molto grandi. Si tratta di una affermazione piuttosto generica che si può precisare in un certo numero di modi alternativi, uno dei quali è il teorema di Chebyshev.

Teorema di Chebyshev: Sia x una variabile casuale tale che esistano finiti la media μ e la varianza σ^2 ; preso un numero reale positivo $c > 0$ si ha:

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

La disuguaglianza di Chebyshev invece dice che

$$P(m - c\sigma \leq x_i \leq m + c\sigma) \geq 1 - \frac{1}{c^2}$$

ovvero se io non conosco la forma di distribuzione ma invece conosco σ e la media posso calcolare la probabilità di un intervallo simmetrico composto da x_i facendo $1 - 1/(\text{il numero di deviazioni dalla media})^2$. Ad esempio: 50 ragazzi ($n=50$) fanno un esame con media voto 24 e deviazione 2, quale è la percentuale di persone tra 21 e 27? dato che $21-24=3$ e $\sigma = 2$ saremo a $3/2$ deviazioni dalla media, quindi

$$1 - \frac{1}{(\frac{3}{2})^2} = 0.56$$

allora sappiamo che la percentuale di persone che hanno ricevuto un voto tra 21 e 27 è non superiore al 56%

momenti di una distribuzione

Definiamo il momento di ordine n di una variabile casuale x attorno ad un punto x_0 come:

$$M_n(x_0) = E[(x - x_0)^n] \begin{cases} \sum_k (x_k - x_0)^n P(x_k) & \text{per variabili discrete} \\ \int_{-\infty}^{+\infty} (x - x_0)^n P(x) dx & \text{per variabili continue} \end{cases}$$

Hanno particolare rilevanza i *momenti algebrici* λ_n , ossia i momenti di ordine generico attorno al punto $x_0 = 0$

$$\lambda_n = M_n(0)$$

ed i *momenti centrali* μ_n , ossia i momenti di ordine generico attorno al valor medio μ di x :

$$\mu_n = M_n(\mu)$$

In particolare la media è il momento algebrico di ordine uno mentre la varianza è il momento centrale di ordine 2. Quello di ordine tre è il primo momento centrale di ordine dispari che non si annulla ed è importante perché, pesando con il segno le code a destra e sinistra della media, misura l'eventuale asimmetria della funzione di distribuzione. Sfruttando la linearità del valore di aspettazione, il momento centrale di ordine tre si può scrivere come

$$\mu_3 = E[x^3] - 3\mu\sigma^2 - \mu^3$$

A partire dal momento centrale di ordine tre si può definire il *coefficiente di asimmetria o skewness* γ_1

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

La skewness è una quantità adimensionale che vale zero per le distribuzioni simmetriche rispetto al valor medio (non è in generale vero il viceversa) e che è diversa da zero se la funzione di distribuzione presenta una coda più lunga dell'altra—nel qual caso una skewness positiva (negativa) indica che la coda a destra (sinistra) è più lunga.

La funzione cumulativa

Data una variabile casuale x la funzione cumulativa è definita come

$$F(x') = P(x < x')$$

La funzione cumulativa è una funzione che leggendo sul dominio da sinistra a destra ad ogni punto della funzione somma la probabilità di quel punto a tutti i precedenti, infatti ha come dominio lo stesso del grafico di distribuzione, come minimo ha 0 e come massimo ha 1 ed essendo che la probabilità non può essere negativa è crescente e monotona quindi i limiti all' ∞ sono sempre 0 e 1

Cenni alle variabili multivariate: indipendenza, covarianza e correlazione

La nostra discussione, fino a questo momento, è stata largamente incentrata su variabili casuali singole, per cui completiamo il capitolo colmando questa lacuna.

L'indipendenza tra variabili casuali

Per semplicità, ma senza perdere di generalità, in questa sezione ci concentreremo sul caso di due variabili casuali continue x_1 ed x_2 , descritte dalla densità di probabilità congiunta $P(x_1, x_2)$. Analogamente a quanto abbiamo fatto nel caso di una variabile, richiederemo che la densità di probabilità sia positiva e propriamente normalizzata sul suo supporto di definizione.

$$p(x_1, x_2) \geq 0 \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 dx_2$$

a questo punto si possono rivisitare tutte le formule già viste sostituendo la formula ad una variabile con quella a due, il caso più interessante sarà quello della probabilità condizionata, ovvero

$$P(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)}$$

$$E[x_1 x_2] = E[x_1] E[x_2]$$

Covarianza e correlazione

Date due variabili casuali x_1 ed x_2 , e dette $\mu_1 = E[x_1]$ e $\mu_2 = E[x_2]$ le rispettive medie, definiamo la loro covarianza, che chiameremo $Cov(x_1, x_2)$, come il valore di aspettazione delle rispettive fluttuazioni attorno al valor medio

$$Cov(x_1, x_2) = E[(x_1 - \mu_1)(x_2 - \mu_2)] = E(x_1 x_2) - E(x_1)E(x_2)$$

Si dimostra inoltre che

$$Cov(x_1, x_2) = E[x_1 x_2] - E[x_1]E[x_2] \rightarrow \text{ se } E(x_1 x_2) = E(x_1)E(x_2) \text{ allora } Cov = 0$$

e che quindi se due variabili sono indipendenti, la loro covarianza è nulla. Non è tuttavia detto il contrario, ovvero il fatto che la covarianza tra due variabili sia nulla non implica necessariamente che le due variabili siano indipendenti e c'è la possibilità che due variabili dipendenti abbiano $E(x_1 x_2) = E(x_1)E(x_2)$. Il motivo è che la covarianza è sensibile solo ad un tipo specifico di mutua dipendenza, ovvero quello lineare. La covarianza è un'applicazione bilineare simmetrica e gode delle classiche sue proprietà. Si dimostra anche che la covarianza di una variabile con se stessa $Cov(x, x)$ non è altro che la varianza e che la covarianza di una variabile con una costante $Cov(c, x)$ vale 0. Essendo bilineare simmetrica è possibile farne la matrice

$$\Sigma = \begin{vmatrix} Cov(x_1 x_1) & Cov(x_1 x_2) \\ Cov(x_2 x_1) & Cov(x_2 x_2) \end{vmatrix} = \begin{vmatrix} \sigma_1^2 & Cov(x_1 x_2) \\ Cov(x_2 x_1) & \sigma_2^2 \end{vmatrix}$$

che può essere semplificato sapendo che per variabili uguali la covarianza è uguale alla varianza, Possiamo definire la correlazione come

$$Corr(x_1, x_2) = \rho_{x_1 x_2} = \frac{Cov(x_1, x_2)}{\sigma_1 \sigma_2}$$

La correlazione possiede banalmente tutte le proprietà della covarianza, in particolare è nulla se le variabili x_1 ed x_2 sono indipendenti (ma, di nuovo, l'implicazione inversa non è vera in generale). La correlazione è una quantità adimensionale che assume valori compresi tra -1 ed 1 (non lo dimostreremo formalmente, ma la covarianza è una sorta di prodotto scalare, mentre la deviazione standard ha il significato di una norma, per cui questa proprietà segue essenzialmente dalla disuguaglianza di Cauchy-Schwarz). Se la correlazione tra due variabili è nulla, esse si dicono scorrelate. Se la correlazione è maggiore (minore) di 0, esse si dicono positivamente (negativamente) correlate. Nel caso particolare in cui una variabile casuale dipenda linearmente da un'altra

$$x_2 = m x_1 + q \text{ per cui } (\sigma_2 = |m| \sigma_1)$$

allora si ha

$$Corr(x_1, x_2) = \frac{Cov(x_1, m x_1 + q)}{\sigma_1 \sigma_2} = \frac{m Cov(x_1, x_1)}{|m| \sigma_2^2} = \frac{m}{|m|} = \pm 1$$

ovvero se due variabili sono dipendenti hanno correlazione ± 1 ; ne segue banalmente che la correlazione di una variabile con se stessa vale 1 ed è possibile costruire una matrice anche per la correlazione

$$R = \begin{vmatrix} Corr(x_1 x_1) & Corr(x_1 x_2) \\ Corr(x_2 x_1) & Corr(x_2 x_2) \end{vmatrix} = \begin{vmatrix} 1 & Corr(x_1 x_2) \\ Corr(x_2 x_1) & 1 \end{vmatrix}$$

In breve

- La probabilità è una misura che associa ad ogni elemento dello spazio degli eventi un numero reale che soddisfa gli assiomi di Kolmogorov.
- La probabilità condizionata $P(E_1|E_2)$ è la probabilità che si verifichi E_1 , subordinata al fatto che si sia verificato E_2 . I due eventi E_1 ed E_2 si dicono indipendenti se $P(E_1|E_2) = P(E_1)$ e $P(E_2|E_1) = P(E_2)$. La probabilità che due eventi indipendenti si verifichino insieme è uguale al prodotto delle probabilità.
- Il teorema di Bayes lega tra di loro le probabilità condizionate $P(E_1|E_2)$ e $P(E_2|E_1)$ ed è di fondamentale importanza nei problemi di probabilità inversa.
- Data una variabile casuale x , la funzione di distribuzione è una funzione che associa ad ogni valore di x la sua probabilità (nel caso discreto) o la sua densità di probabilità (nel caso continuo).

- A partire dal concetto di valore di aspettazione di una funzione di variabile casuale si definiscono tutte le proprietà rilevanti delle distribuzioni, tra cui

$$\mu = E[x] \quad (\text{la media})$$

$$\sigma^2 = E[(x - \mu)^2] = E[x^2] - \mu^2 \quad (\text{la varianza})$$

$$\sigma = \sqrt{\sigma^2} \quad (\text{la deviazione standard})$$

$$\mathbf{M}_n(x_0) = E[(x - x_0)^n] \quad (\text{i momenti di ordine generico})$$

Capitolo 4

Variabili campione e propagazione dell'errore statistico

Quando misuriamo una grandezza fisica (e specialmente quando il valore della misura fluttua) possiamo pensare che il valore stesso sia una variabile casuale con una certa distribuzione (tipicamente incognita a priori) che chiamiamo *distribuzione generatrice*. In questo schema concettuale fare n misure di una stessa grandezza fisica in condizioni di ripetitività equivale a campionare n volte la distribuzione generatrice corrispondente. È ovvio, allora, che in generale non potremo conoscere mai perfettamente la forma della distribuzione generatrice, né le sue caratteristiche (e.g., la media o la deviazione standard) ma è altresì ovvio che, in questo processo di campionamento, **acquisiamo progressivamente informazioni sulla distribuzione stessa**. Ritornando al problema iniziale che avevamo con l'errore massimo allora possiamo riformulare il tutto con questa forma di errore e di misurazioni, dove la miglior stima sarà la media e l'errore sarà la deviazione standard, indicheremo il tutto così:

$$x = \tilde{x} \pm \sigma_x [\text{unità di misura}]$$

È chiaro che questa nuova formula differisce non solo dall'errore massimo per l'aggiunta di informazione ma anche poiché non da un intervallo certo ma da invece un intervallo in cui c'è una determinata probabilità che l'evento accada, chiameremo questa probabilità confidence level [CL] e sarà utile specificarla sempre.

Campionamenti singoli

Se è conosciuto il grafico di distribuzione o la deviazione standard a priori possiamo prendere una sola misura ed utilizzarla come miglior stima e porre la deviazione standard come errore, possiamo anche decidere che grado di confidenza vogliamo per poi mettere un multiplo o sottomultiplo della deviazione affinché si raggiunga quel grado di confidenza. Se la deviazione standard della distribuzione generatrice non è nota a priori, possiamo stimarla a posteriori in base ad una serie di misure ripetute, da un punto di vista operativo il nostro problema fondamentale diventa: data una serie di n misure di una stessa grandezza fatte in condizioni di ripetitività, quali sono le migliori stime che possiamo dare della media e della deviazione standard della distribuzione generatrice?

Somma di variabili casuali indipendenti

Date due variabili casuali indipendenti x_1 ed x_2 (continue o discrete) caratterizzate dalle rispettive funzioni di distribuzioni (in generale diverse), la loro somma $x = x_1 + x_2$ è ovviamente ancora una variabile casuale, è allora lecito chiedersi quale sia la funzione di distribuzione di x e come essa si possa calcolare a partire dalle funzioni di distribuzione di x_1 ed x_2 .

Media e varianza

La media μ molto semplicemente è

$$\mu = E[x_1 + x_2] = E[x_1] + E[x_2] = \mu_1 + \mu_2$$

mentre la varianza σ^2 diventa dopo alcuni passaggi

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2Cov(x_1, x_2)$$

dove Cov indica la covarianza che nelle variabili indipendenti è per definizione nulla, quindi varianza e deviazione standard si semplificano come

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \text{ da cui } \sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$$

inifine è possibile generalizzare il tutto così

$$\text{media} = \Sigma_{i=1}^n \mu_i \quad \text{varianza} = \Sigma_{i=1}^n \sigma_i^2 \quad \text{deviazione standard} = \sqrt{\Sigma_{i=1}^n \sigma_i^2}$$

Misure ripetute: media e varianza campione

La media campione m non è altro che la media aritmetica delle misure ed il suo valore di aspettazione è uguale alla media μ (in generale un elemento con questa proprietà viene chiamato imparziale). Differisce da quest'ultima perché è una stima e non ne abbiamo i valori esatti

$$m = \frac{1}{n} \sum_{i=1}^n x_i$$

La varianza invece è leggermente più complicata, se conoscessimo a priori la media potremmo fare

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

il cui valore di aspettazione, come prima, coincide con la quantità che vogliamo stimare (ed è quindi imparziale) infatti

$$E[s^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right] = \frac{1}{n} \sum_{i=1}^n E\left[(x_i - \mu)^2\right] \frac{1}{n} \sum_{i=1}^n \sigma^2 = \sigma^2$$

il problema a questo punto sta nel fatto che noi non conosciamo la media ma solo una sua stima. Potremmo allora prendere

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$$

con un calcolo della derivata si può vedere come questa sia in realtà una sottostima di σ^2 e sia minimizzato quando una generica media è uguale alla media campione. Il valore di aspettazione di s^2 si calcola come

$$E[s_n^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - m)^2\right] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i^2 + m^2 + 2mx_i)\right]$$

e poiché gli x_i hanno distribuzione generatrice uguale è possibile calcolarne un pezzo per volta

$$E[x_i^2] = \sigma^2 + \mu^2$$

$$E[m^2] = \frac{1}{n} \sigma^2 + \mu^2$$

$$E[mx_i] = \frac{1}{n} \sigma^2 + \mu^2$$

Quindi nel totale

$$E[s_n^2] = \frac{n-1}{n} \sigma^2$$

La varianza campione, scritta come all' inizio , non è imparziale se non asintoticamente, nel senso che il suo valore di aspettazione, in generale, non è uguale alla varianza della distribuzione generatrice se non per $n \rightarrow \infty$. Ovviamente non sono tutti valori indipendenti poiché dipendono da m , se infatti conoscessimo m e $n-1$ valori si potrebbe risalire all' ultimo. Si può ottenere una stima imparziale della varianza della distribuzione generatrice semplicemente moltiplicando per il fattore correttivo $n/n-1$

$$s_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \text{ che soddisfa banalmente } E[s_{n-1}^2] = \sigma^2$$

La deviazione standard della media

Come adattiamo adesso queste stime effettuate alla nuova forma di scrittura dell'errore statistico? Ovviamente la media campione sarà la stima centrale, ma per l'incertezza come facciamo? Potremmo prendere la varianza campione ma essa si basa sulle fluttuazioni delle singole misure mentre noi vogliamo solo le fluttuazioni del valore medio. Prendiamo quindi $\text{Var}(m)$ che possiamo calcolare così

$$\text{Var}(m) = \text{Var}\left(\frac{1}{n} \sum_{x=1}^n x_1\right) = \frac{\sigma^2}{n}$$

da cui la deviazione standard della media è

$$\sigma_m = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

La precedente formula ci indica come le fluttuazioni della media di n misure si riducono, al variare di n , rispetto alle fluttuazioni sulla singola misura. Possiamo utilizzare questo risultato, insieme alla stima della varianza campione, per la stima della varianza e della deviazione standard della media s_m

$$s_m^2 = \frac{1}{n(n-1)} \sum_{i=1}^n (x_i - m)^2 \text{ e } s_m = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - m)^2}$$

Covarianza e correlazione campione

Supponiamo di avere una serie di campionamenti x_i e y_i , possiamo analogamente alla media e varianza campione stimare la covarianza q e correlazione campione così

$$q_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_x)(y_i - m_y)$$

e a questo punto possiamo trovare la correlazione r dividendo covarianza per deviazione standard di x e y

$$r_{xy} = \frac{q_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_{i=1}^n (x_i - m_x)^2 \sum_{i=1}^n (y_i - m_y)^2}}$$

Uno dei modi più semplici per individuare una possibile correlazione tra due variabili in un campione è quello di realizzare un grafico di dispersione. Se valori piccoli di una variabile tendono ad essere associati a valori piccoli (o grandi) dell'altra, e viceversa, questo è indice di una correlazione positiva (o negativa). Se le fluttuazioni attorno al valor medio delle due quantità appaiono indipendenti, cioè la distribuzione dei punti nel diagramma di dispersione non è elongata in una direzione privilegiata, allora le quantità sono scorrelate. L'indice di Pearson indica l'entità della correlazione ed è ± 1 se e solo se le due grandezze nel grafico di dispersione stanno esattamente su una retta. Per quanto la tentazione di fare il collegamento possa essere forte, non si può sottolineare abbastanza il fatto che una correlazione tra due grandezze non implica necessariamente un rapporto di causa-effetto tra le grandezze stesse.

Media e varianza di una funzione di variabili casuali

Il caso unidimensionale

Consideriamo dunque una variabile casuale x , di cui assumiamo di sapere la funzione di distribuzione, sia μ il valor medio di x e σ la sua deviazione standard, il nostroscopto è quello di stimare la media μ_f e la varianza σ_f^2 di una generica funzione $f(x)$. Poiché sappiamo che i valori di x tendono a concentrarsi attorno al valor medio, possiamo partire dallo sviluppo in serie di Taylor di $f(x)$ attorno a μ troncato al prim'ordine:

$$f(x) \approx f(\mu) + \frac{df}{dx}(\mu)(x - \mu) \text{ ovvero } f(x) - f(\mu) \approx \frac{df}{dx}(\mu)(x - \mu)$$

La media si calcola formalmente come il valore di aspettazione $E[f(x)]$, che sviluppato al prim'ordine diviene

$$\mu_f = E[f(x)] = E\left[f(\mu) + \frac{df}{dx}(\mu)(x - \mu)\right] = E[f(\mu)] + \frac{df}{dx}(\mu)E[(x - \mu)]$$

(Ricordate: $f(\mu)$ è una costante che può essere portata fuori dal valore di aspettazione, e così pure la derivata, sono entrambe calcolate in un punto e dunque non dipendono da x .) A questo punto, notando che $E[x - \mu]$ si annulla banalmente per la linearità del valore di aspettazione, si ottiene

$$\mu_f \approx f(\mu)$$

Il valor medio di $f(x)$, può essere approssimato con il valore della funzione $f(x)$ calcolata nel valor medio di x , a meno di termini del second'ordine in $(x - \mu)$. Che il termine al prim'ordine dello sviluppo si annulli è un fatto importante, che ci dice che la precedente equazione vale esattamente (senza resto) se $f(x)$ è una funzione lineare di x . Per la varianza possiamo procedere analogamente, utilizzando il nostro sviluppo in serie di $f(x)$ ed il risultato che abbiamo appena ottenuto per la media

$$\sigma_f^2 = E[(f(x) - \mu_f)^2] \approx E[(f(x) - f(\mu))^2] \approx E\left[\left(\frac{df}{dx}(\mu)(x - \mu)\right)^2\right] \approx \left(\frac{df}{dx}(\mu)\right)^2 E[(x - \mu)^2]$$

Il caso generale

Nel caso multidimensionale (quindi $f(x_1, x_2, x_3 \dots)$ con $x_1, x_2, x_3 \dots$ variabili $\mu_1, \mu_2, \mu_3 \dots$ medie e $\sigma_1^2, \sigma_2^2, \sigma_3^2 \dots$ varianze) la situazione si complica a livello di calcoli ma sulla teoria non cambia, iniziando da un caso a due variabili, prima si effettua lo sviluppo di Taylor al primo ordine

$$f(x_1, x_2) \approx f(\mu_1, \mu_2) + \frac{\partial f}{\partial x_1}(\mu_1, \mu_2)(x_1 - \mu_1) + \frac{\partial f}{\partial x_2}(\mu_1, \mu_2)(x_2 - \mu_2)$$

seguendo i passaggi effettuati nel caso unidimensionale troviamo la media

$$\mu_f = E[f(x_1, x_2)] \approx f(\mu_1, \mu_2) + \frac{\partial f}{\partial x_1}(\mu_1, \mu_2)E[x_1 - \mu_1] + \frac{\partial f}{\partial x_2}(\mu_1, \mu_2)E[x_2 - \mu_2] = f(\mu_1, \mu_2)$$

e la varianza

$$\begin{aligned} \sigma_f^2 &= E[(f(x_1, x_2) - f(\mu_1, \mu_2))^2] \approx (\text{dopo alcuni calcoli}) \\ &= \left(\frac{\partial f}{\partial x_1}(\mu_1, \mu_2)\right)^2 \sigma_1^2 + \left(\frac{\partial f}{\partial x_2}(\mu_1, \mu_2)\right)^2 \sigma_2^2 + 2 \frac{\partial f}{\partial x_1}(\mu_1, \mu_2) \frac{\partial f}{\partial x_2}(\mu_1, \mu_2) \text{Cov}(x_1, x_2) \end{aligned}$$

è possibile sostituire Cov con Corr $\sigma_1 \sigma_2$. Inoltre per il caso a più dimensioni la soluzione è ovvia (altrimenti statnotes p.103)

Propagazione dell'errore statistico

In poche parole se le variabili sono indipendenti la formula è

$$\sigma_f^2 \approx \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(\tilde{x}_1 \dots \tilde{x}_n)\right)^2 \sigma_i^2 \quad \text{ovvero} \quad \sigma_f \approx \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(\tilde{x}_1 \dots \tilde{x}_n)\right)^2 \sigma_i^2}$$

Il contenuto fisico fondamentale è che nel caso di grandezze indipendenti (che è tipico in laboratorio) gli errori statistici non si sommano linearmente come gli errori massimi, ma in quadratura. Possiamo andare oltre, se le grandezze non sono indipendenti ed abbiamo delle stime r_{ij} dei coefficienti di correlazione $\text{Corr}(x_i, x_j)$ la precedente equazione si generalizza come

$$\sigma_f^2 \approx \sum_{i=1}^n \sum_{j=1}^n \frac{\partial f}{\partial x_i}(\tilde{x}_1 \dots \tilde{x}_n) \frac{\partial f}{\partial x_j}(\tilde{x}_1 \dots \tilde{x}_n) r_{ij} \sigma_i \sigma_j$$

Capitolo 5

Distribuzioni univariate di uso comune

La distribuzione binomiale

Prima di parlare di distribuzione binomiale parliamo del diagramma delle probabilità a due uscite (Il diagramma ad albero utilizzato per vedere la probabilità che un evento dei due possibili ad ogni scelta accada). In formule: dati due eventi $P(E_1)=p$ e $P(E_2)=1-p$ la probabilità che E_1 accada k volte è

$$P(E_1^1 \dots E_1^k, E_2^1 \dots E_2^{n-k}) = p^k (1-p)^{n-k}$$

questa formula contiene, tuttavia, sola delle varie combinazioni possibili. Per calcolare il numero di possibili combinazioni e la probabilità di quel particolare evento si realizzi, altri non è che la distribuzione del binomio

$$B(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

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

dove il binomio ci da il numero di volte che un determinato evento accade mentre la precedente formula ci da la probabilità

Normalizzazione, media e varianza

Risulta ovvio e di facile calcolo che $\sum B(k; n, p) = 1$, questa è chiamata "normalizzazione di una distribuzione" e serve a dimostrare che la somma delle varie probabilità che gli eventi accadano sia 1 (come è logico che debba essere). Anche la media secondo la definizione diventa

$$\mu = E[k] = \sum k B(k; n, p) = np$$

la varianza diventa

$$\sigma^2 = E[k^2] = \sum k^2 B(k; n, p) = np(1-p)$$

(guardare esempio 5.2 p.111)

momenti di ordine superiore

per i momenti di ordine superiore il calcolo risulta tedioso nonostante utilizzi sempre le solite formule, quindi dopo vari calcoli riformuliamo la formula dei momenti così

$$E[k^{m+1}] = p(1-p) \frac{d}{dp} E[k^m] + np E[k^m]$$

da cui possiamo ricavare $E[k^3]$ e l'indice di asimmetria

$$E[k^3] = np(1-p)(1-2p) + 3n^2 p^2 (1-p) + n^3 p^3$$

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{(1-2p)}{\sqrt{np(1-p)}}$$

La distribuzione multinomiale

La distribuzione multinomiale è una generalizzazione della distribuzione binomiale in cui lo schema è ancora quello delle n ripetizioni indipendenti di uno stesso esperimento, ma stavolta l'esperimento in questione può avere m (anziché 2) esiti distinti e mutuamente esclusivi $E_1 \dots E_m$ con probabilità associate $p_1 \dots p_m$ soggette agli usuali vincoli

$$p_1 \geq 0 \quad \sum_k^m p_k = 1$$

Ripassando le proprietà e formule della distribuzione binomiale risulta:

La probabilità che su n ripetizioni l'evento E_1 si verifichi k_1 volte, l'evento E_2 si verifichi k_2 volte ... e l'evento E_m si verifichi k_m volte è

$$M(k_1 \dots k_m; n, p_1 \dots p_m) = \binom{n}{k_1 \dots k_m} p_1^{k_1} \dots p_m^{k_m} = \frac{n!}{k_1! \dots k_m!} p_1^{k_1} \dots p_m^{k_m}$$

La media e la varianza sono corrispettivamente

$$E[k_i] = np_i \quad Var(k_i) = np_i(1 - p_i)$$

Una cosa che possiamo chiederci a questo punto è se i k_i , visti come variabili indipendenti, siano tra loro indipendenti. Non sorprendentemente, la risposta è no, perché, come abbiamo detto prima, essi sono legati dalla relazione

$$\sum_{i=1}^m k_i = n$$

Fisicamente questo corrisponde a dire che se lanciamo 100 volte un dado e si osserva, ad esempio, un numero di uscite della faccia 2 superiore alla media, questo dovrà essere bilanciato da un numero di uscite delle altre facce (alcune o tutte) inferiore alla media. (Per inciso, questo semplice argomento ci dice che la correlazione tra facce diverse deve essere negativa.) In effetti si può dimostrare che

$$Cov(k_i, k_j) = -np_i p_j \quad \text{da cui} \quad Corr(k_i, k_j) = -\sqrt{\frac{p_i p_j}{(1 - p_i)(1 - p_j)}}$$

La distribuzione di poisson

Consideriamo un processo stazionario nel dominio del tempo che produce il verificarsi di eventi indipendenti l'uno dall'altro, l'evento elementare potrebbe essere, ad esempio, il passaggio di una macchina da una strada trafficata, o l'arrivo di un cliente in un ufficio postale, o l'incontro di un conoscente in un luogo affollato. Le due domande con cui apriamo questa sezione sono:

- cosa possiamo dire sulla probabilità che, in un dato intervallo di tempo, si verifichino k eventi quando in media se ne verificano μ ?
- tutto questo ha qualcosa a che vedere con la distribuzione binomiale che abbiamo appena analizzato in dettaglio?

Processi Poissoniani

Quando parliamo di processo Poissoniano (e, per fissare le idee, di processo Poissoniano nel dominio del tempo) intendiamo essenzialmente tre cose:

- **Indipendenza** Ogni evento elementare è indipendente dagli altri ovvero il verificarsi o non verificarsi di tale evento non influenza gli altri eventi.
- **Stazionarietà** La frequenza temporale media degli eventi è indipendente dal tempo, cioè il numero medio di eventi per unità di tempo (*probabilità media in un intervallo che un evento accada*) λ è lo stesso in qualsiasi intervallo.
- **Non simultaneità** Non si possono verificare due o più eventi elementari nello stesso istante.

Consideriamo dunque un intervallo di tempo Δt , che dividiamo in n intervalli più piccoli di lunghezza $dt = \frac{\Delta t}{n}$ con l'idea che alla fine $n \rightarrow \infty$ così che $\mu = \lambda \Delta t$ è il numero medio di eventi nell'intervallo di tempo (λ è il numero di eventi per unità di tempo dt) (dividiamo in entrambi i lati per avere un sottointervallo oppure anche dt). A questo punto se riusciamo ad ottenere un intervallo abbastanza piccolo come dt possiamo considerare la terza proprietà per notare come in quel intervallo il numero di eventi può essere 1 (con probabilità p) o 0 (con probabilità $(1-p)$) e scrivere che il numero medio di intervalli è

$$\lambda dt = \frac{\mu}{n} = 0 \times (1-p) + 1 \times p = p$$

ciò vuol dire che λdt non è soltanto il numero medio di eventi ma anche la proprietà che si verifichi il singolo evento. Questo ci offre la possibilità di calcolare immediatamente un certo numero di cose interessanti, ad esempio: qual è la probabilità $P(0)$ di osservare 0 eventi nell'intero intervallo t quando, come abbiamo visto, in media ne osserviamo μ ? Beh, visto che per ipotesi gli eventi sono indipendenti, le probabilità corrispondenti si moltiplicano e nel limite $n \rightarrow \infty$ la risposta è

$$P(0; \mu) = \lim_{n \rightarrow \infty} (1-p)^n = \lim_{n \rightarrow \infty} \left(1 - \frac{\mu}{n}\right)^n = e^{-\mu}$$

E se invece ci chiedessimo qual è la probabilità di osservare esattamente 1 evento nell'intero intervallo t ? Il ragionamento procede essenzialmente come prima, salvo il fatto che nel prodotto dobbiamo sostituire con p uno degli n termini $(1-p)$ (stiamo richiedendo che non si verifichino eventi in $n-1$ intervallini e che si verifichi un evento nell'intervallino rimanente) e dobbiamo moltiplicare per n , per tenere conto del fatto che siamo liberi di scegliere a caso tra gli n disponibili

$$P(1; \mu) = \lim_{n \rightarrow \infty} np(1-p)^{n-1} = \lim_{n \rightarrow \infty} \frac{n\mu}{n} \left(1 - \frac{\mu}{n}\right)^{n-1} = \mu e^{-\mu}$$

Stiamo di fatto riderivando la distribuzione binomiale nel limite in cui la probabilità del singolo evento $p = \lambda dt = \mu/n$ è piccola ed il numero di ripetizioni n è grande. Ma allora abbiamo tutti gli strumenti per calcolare la forma esplicita della funzione di distribuzione che chiameremo distribuzione di Poisson.

La distribuzione di Poisson come limite della binomiale

Come abbiamo visto la distribuzione di Poisson si ottiene come limite della distribuzione binomiale quando $n \rightarrow \infty$ e $p \rightarrow 0$ in modo che la media np si mantenga costante, notiamo allora che $(1-p) \rightarrow 1$ e $\sigma^2 = np(1-p) = np = \mu$ osserviamo adesso che se utilizzo il limite della distribuzione binomiale e aggiungo ad essa le proprietà della distribuzione poissoniana ottengo

$$\lim_{n \rightarrow \infty} B(k; n, p) = \frac{n!}{k!(n-k)!} \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} = \frac{n(n-1) \dots (n-k+1)}{n^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{n}\right)^{n-k}$$

Attenzione k non è un numero ma una variabile casuale che assume valori da 0 a ∞ poiché non siamo più nel caso 1 o 0 il che non rende chiara questa formula, allora introduciamo una nuova variabile

$$\xi = \frac{k - np}{n}$$

il numeratore misura una sorta di deviazione standard ed è quindi dello stesso ordine di $\sigma = \sqrt{np(1-p)}$ allora $\xi \rightarrow 0$ come $\frac{1}{\sqrt{n}}$ per $n \rightarrow \infty$ adesso tramite alcune sostituzioni possiamo arrivare alla seguente espressione

$$P(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}$$

che è l'espressione chiusa della distribuzione di Poisson e possiamo confermare che è un caso limite della distribuzione binomiale. Inoltre alla distribuzione binomiale dipende da due parametri, n e p , che concorrono entrambi a determinare il valor medio e la varianza; la distribuzione di Poisson dipende da un solo parametro, la media che incidentalmente coincide anche con la varianza. Ma la cosa più importante è che, nello schema binomiale, la variabile casuale k è limitata superiormente ($k \leq n$), mentre in quello di Poisson k può assumere qualsiasi valore intero da 0 a ∞ , anche se in pratica la probabilità corrispondente decresce velocemente al crescere di k . Quando vi chiedete se un certo fenomeno segua la statistica binomiale o Poissoniana, questo è un buon indicatore: il valore massimo di occorrenze è fissato oppure no?

Normalizzazione, media e varianza

Come già visto la media vale $np \rightarrow \mu$ e la varianza $np(1-p) \rightarrow \mu$ ma rendiamolo esplicito

$e^\mu e^{-\mu} = 1$ e quindi risulta correttamente normalizzata

$$E[k] = \sum_{k=0}^{\infty} k P(k; \mu) = \sum k \frac{\mu^k}{k!} e^{-\mu}$$

pongo $h = k - 1$ e risulta

$$\mu e^\mu e^{-\mu} = \mu$$

Per la varianza facciamo uguale

$$E[k^2] = \mu^2 + \mu \text{ quindi } \sigma^2 = \mu$$

Momenti di ordine superiore

Esattamente come abbiamo visto per la distribuzione binomiale, anche per la distribuzione di Poisson è possibile ricavare una relazione ricorsiva che consente di calcolare i momenti algebrici di ordine generico a partire da quelli di ordine più basso. Il ragionamento procede in modo analogo

$$\frac{d}{d\mu} E[k^m] \rightarrow E[k^{m+1}] \mu \left(\frac{d}{d\mu} E[k^m] + E[k^m] \right)$$

si ricava così il momento di terzo grado che vale μ e così $\gamma_1 = \frac{1}{\sqrt{\mu}}$

Distribuzione di Poisson e distanza tra eventi successivi

Come si calcola la distanza temporale media tra due eventi? Ovviamente si tratta di una variabile aleatoria altrimenti sarebbero fissi anche il numero k di eventi, il calcolo è tuttavia semplice infatti

$$E[t] = \frac{\Delta t}{\mu} = \frac{\Delta t}{\Delta t \lambda} = \frac{1}{\lambda}$$

essendo λ la frequenza con quindi unità di misura t^{-1} il calcolo è dimensionalmente corretto. Inoltre possiamo calcolare l'espressione esplicita per la funzione di distribuzione di t notando che, dato il fatto che un evento si sia verificato ad un certo istante t_0 , la probabilità (infinitesima) che l'evento successivo si verifichi entro un intervallino (infinitesimo) di durata dt ad una distanza temporale t è data dal prodotto

$$dP(t, dt) = e^{-\lambda t} \times \lambda dt$$

in cui il primo termine rappresenta la probabilità che non si verifichi nessun evento tra i tempi t_0 e $t_0 + t$ ed il secondo la probabilità che si verifichi esattamente un evento tra i tempi $t_0 + t$ e $t_0 + t + dt$. Dividendo entrambi i membri per dt possiamo calcolare la probabilità specifica per unità di tempo, vale a dire la nostra densità di probabilità.

$$P(t, \lambda) = \lambda e^{-\lambda t}$$

Tutto ciò fa notare come le distanze tra gli eventi di una distribuzione poissoniana siano esponenziali (lo studieremo più avanti).

Somma di variabili Poissoniane

La somma di due distribuzioni poissoniane $k=l+m$ si scrive nel seguente modo:

$$P(k) = \sum_{l=0}^k P(l; \mu_1) P(k-l; \mu_m) = P(K; \mu_1 + \mu_m)$$

La distribuzione uniforme

La distribuzione uniforme è l'esempio più semplice di funzione di distribuzione di variabile casuale continua, la densità di probabilità è costante entro un intervallo finito e nulla fuori:

$$u(x; a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x \leq a \text{ o } x \geq b \end{cases}$$

Normalizzazione, media, varianza e coefficiente di asimmetria

Si dimostra che è correttamente normalizzata

$$\int_{-\infty}^{\infty} u(x; a, b) dx = \int_b^a \frac{1}{b-a} dx = \frac{1}{b-a} (b-a) = 1$$

La media è data facilmente secondo la formula del valore di aspettazione come

$$\frac{b+a}{2}$$

uguale per il momento di secondo grado $E[x^2]$ che fa $\frac{b^2+ab+a^2}{3}$ e di conseguenza la deviazione standard che fa

$$\sigma = \frac{b-a}{\sqrt{12}}$$

Questa $\sqrt{12}$ al denominatore della deviazione standard è un numero importante da tenere a memoria, perché è quello che determina l'incertezza di misura (nel senso statistico) di tutti gli strumenti digitali— cioè di quegli strumenti che forniscono in uscita valori discreti (spaziati tra di loro di una quantità pari alla risoluzione strumentale) non affetti da fluttuazioni statistiche. Per completezza possiamo rivangare il concetto di semilarghezza che sarà ovviamente la metà del intervallo

$$HWHM = \frac{b-a}{2} = 1,73\sigma$$

comulativa e quantili

La funzione cumulativa è semolicemnete

$$F(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a} \quad a \leq x \leq b$$

La funzione di distribuzione inversa (quantili) è invece

$$F^{-1}(u) = q(b-a) + a$$

La distribuzione esponenziale

Avevo già annunciato qualcosa sull' esponenziale con la distanza tra gli eventi. In generale, dato un numero positivo $\lambda > 0$ una distribuzione della forma

$$\xi(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x < \infty \\ 0 & x \leq a \end{cases}$$

si dice esponenziale con parametro λ . Vedremo tra un attimo che λ (che per ragioni dimensionali ha le dimensioni di x^{-1}) ha il significato fisico dell'inverso della media della distribuzione, per cui a volte l' espressione si trova anche scritta in funzione di λ^{-1} che prende il nome di vita media nel caso di processi Poissoniani nel dominio del tempo e cammino libero medio nel caso di processi Poissoniani nel dominio dello spazio.

Normalizzazione, media, varianza e coefficiente di asimmetria

Come siamo ormai abituati a fare si fanno tutti i calcoli per vedere che è normalizzata

$$\int_0^{\infty} \xi(x; \lambda) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = 1$$

la media sarà uguale alla media della distanza tra gli eventi poissiani

$$\frac{1}{\lambda}$$

la varianza sarà la media alla seconda e la deviazione sarà quindi uguale alla media. Dopo alcuni calcoli troviamo che HWHM è 0.347σ . I momenti algebrici valgono

$$\frac{6}{\lambda} E[x^{n-1}]$$

Funzione cumulativa e quantili

la funzione cumulativa vale

$$F(x) = 1 - e^{-\lambda x}$$

ed i quantili

$$F^{-1}(q) = -\frac{\ln(1-q)}{\lambda}$$

da cui $\gamma_1 = 2$

Assenza di memoria

Una proprietà interessante della distribuzione esponenziale è costituita dal fatto che

$$P(x \geq x_1 + x_2) = \int_{x_1+x_2}^{\infty} \xi(x; \lambda) dx = e^{-\lambda(x_1+x_2)}$$

da cui segue che

$$P(x \geq x_1 + x_2) = P(x \geq x_1)P(x \geq x_2|x \geq x_1)$$

Ora, per definizione di probabilità condizionata, si ha anche che

$$P(x \geq x_1 + x_2) = P(x \geq x_1)P(x \geq x_1 + x_2 | x \geq x_1)$$

Una variabile casuale che goda di questa proprietà si dice una variabile senza memoria

La distribuzione gaussiana

La distribuzione di Gauss può essere vista come il limite di una distribuzione di Poisson per $\mu \rightarrow \infty$ o come il limite di una binomiale con $n \rightarrow \infty$ senza alcuna informazione aggiuntiva su p. Inoltre, per il teorema del limite centrale (studio in seguito), la distribuzione della media di un numero abbastanza grande di campionamenti di una variabile casuale, indipendentemente dalla sua distribuzione, è distribuita Gaussianamente.

La distribuzione di Gauss come limite della Poissoniana

Vogliamo prendere l'espressione della distribuzione di Poisson e farne il limite per μ che tende ad infinito ma troviamo lo stesso problema che abbiamo trovato nel fare il limite della binomiale per trovare la poissoniana, poiché, allo stesso tempo abbiamo bisogno di una prescrizione per k, che è una variabile casuale e, come tale, non è univocamente determinata da μ . Per fare ciò iniziamo facendo il logaritmo di ambo i lati della poissoniana

$$\ln P(k; \mu) = \ln \left(\frac{\mu^k}{k!} e^{-\mu} \right) = k \ln \mu - \ln(k!) - \mu$$

Dopo vari calcoli otteniamo

$$\ln P(k; \mu) \approx -\frac{1}{2} \ln(2\pi\mu) - \frac{1}{2} \frac{(k-\mu)^2}{\mu} \quad \text{ovvero} \quad P(k; \mu) \approx \frac{1}{\sqrt{2\pi\mu}} e^{-\frac{1}{2} \frac{(k-\mu)^2}{\mu}}$$

A questo livello l'espressione della gaussiana è un'espressione approssimata per la poissoniana che può essere utilizzata quando μ è grande. Ma fermiamoci per un attimo ad osservare meglio la relazione che abbiamo appena ricavato: per prima cosa la gaussiana è simmetrica rispetto al valor medio della distribuzione $k = \mu$, questo non dovrebbe sorprendere, poiché sappiamo già che il coefficiente di asimmetria della distribuzione di Poisson vale $\gamma_1 = \frac{1}{\sqrt{\mu}}$, che tende a 0 per $\mu \rightarrow \infty$. (In altre parole sapevamo già dall'espressione poissoniana che la distribuzione di Poisson tende a diventare simmetrica per valori grandi della media.) La seconda osservazione è che, dato che varianza σ^2 della distribuzione di Poisson è uguale al valor medio μ , possiamo in effetti riscrivere la gaussiana nella forma alternativa

$$N(k; \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{k-\mu}{\sigma} \right)^2}$$

Se adesso sostituiamo la variabile casuale discreta k con una variabile continua x, e lasciamo la deviazione standard σ libera di variare indipendentemente dalla media, abbiamo ottenuto la distribuzione di Gauss.

Normalizzazione, media e varianza

La funzione di distribuzione di Gauss (o distribuzione normale) si scrive dunque nella forma

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

dove μ e σ^2 , sono proprio la media e la varianza della distribuzione e sono indipendenti tra di loro (cioè non sono più necessariamente coincidenti come nel caso Poissoniano). Una variabile casuale Gaussiana z con media 0 e varianza 1, che ha funzione di distribuzione (completamente fissata)

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

si dice variabile Gaussiana in forma standard. Una variabile Gaussiana generica x può essere trasformata nella corrispondente variabile in forma standard attraverso il cambio di variabile

$$z = \frac{x - \mu}{\sigma} \quad \text{ovvero} \quad x = z\sigma + \mu$$

Notiamo che il valore massimo della funzione (che si ha per $x=\mu$) è inversamente proporzionale a σ , per cui se raddoppiamo la deviazione standard della distribuzione, stiamo effettivamente raddoppiando la larghezza e dimezzando l'altezza. (L'integrale sotto la curva rimane costante per la condizione di normalizzazione.) La verifica della normalizzazione della forma normale di Gauss si effettua in questo modo

$$\int_{-\infty}^{\infty} N(x; \mu, \sigma) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} I_0$$

Sfortunatamente questo integrale non ha espressione analitica, nel senso che non esiste una primitiva dell'integrando esprimibile in forma chiusa in termini di funzioni elementari (La questione è, almeno in parte, semantica, poiché l'integrale definito di $e^{-\frac{1}{2}z^2}$ tra due estremi arbitrari può essere calcolato numericamente (e tabulato, come vedremo) con il grado di accuratezza desiderato.). Esiste tuttavia un metodo per calcolarlo analiticamente passando alle coordinate polari (metodo di Fubini) ma non lo trascrivo, ad ogni modo la gaussiana risulta correttamente normalizzata. La media la conosciamo già poiché la distribuzione è simmetrica rispetto al punto $x = \mu$ (quindi coincidono anche moda e mediana). La varianza come già anticipato sarà σ^2 . A questo punto gli integrali ci permettono di calcolare banalmente tutti i momenti centrali di ordine superiore, ma dal nostro punto di vista questo non è interessante in quanto il coefficiente di asimmetria γ_1 si annulla per ragioni di simmetria ed i momenti centrali di ordine pari non ci dicono molto altro sulla distribuzione. È invece interessante, come abbiamo fatto con tutte le distribuzioni incontrate sino ad ora, calcolare il coefficiente di proporzionalità tra la semilarghezza a metà altezza e la deviazione standard

$$HWHM = 1.178\sigma$$

L'integrale normale degli errori

Il problema dell'impossibilità di rappresentare quell'integrale nella sua forma indefinita persiste, infatti ciò non ci permette di scrivere la forma analitica della gaussiana e quindi di calcolare la densità di probabilità. Il metodo più ovvio è la funzione cumulativa

$$P(x_1 \leq x \leq x_2) = f(x_2) - f(x_1)$$

Ora, data una variabile Gaussiana arbitraria, il calcolo del valore della funzione cumulativa in un punto generico si può sempre ricondurre all'integrale di una distribuzione di Gauss in forma standard mediante il cambiamento di variabile z per cui in effetti l'unica cosa di cui abbiamo bisogno è una forma tabulata della funzione cumulativa per una variabile Gaussiana in forma standard (1-2-3=68-95-99). Sfortunatamente, per motivi storici, l'integrale che si trova più frequentemente tabulato (ed implementato numericamente nella maggior parte dei linguaggi di programmazione e dei programmi di analisi dati) non è la funzione cumulativa ma la error function, o funzione degli errori definita come

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

È chiaro che le due cose sono legate tra di loro, ma non coincidono esattamente. La funzione degli errori rappresenta la probabilità che una variabile casuale con media 0 e varianza 1/2 cada nell'intervallo $[-x, x]$. La relazione che lega una all'altra è

$$F(x) = \Phi(z) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)$$

che permette di passare dalla error function alla funzione cumulativa della gaussiana. Per una variabile Gaussiana in forma standard z la funzione cumulativa è così importante da meritare un nome tutto per sé (solitamente *phi*). Le appendici A.1-A.5 contengono i valori tabulati di varie quantità legate alla funzione cumulativa di una distribuzione di Gauss in forma standard, e possono essere utili quando non si ha a disposizione un calcolatore.

Il teorema centrale del limite

Abbiamo visto che la distribuzione di Gauss è il limite per $\mu \rightarrow \infty$ della distribuzione di Poisson e, di conseguenza, anche della distribuzione binomiale. Vedremo nel proseguo che altre distribuzioni (e.g., quella del X^2 e la t di Student) tendono ad una Gaussiana nei limiti opportuni. Adesso è il momento di enunciare uno dei risultati più importanti del calcolo delle probabilità, il teorema centrale del limite. **Teorema (centrale del limite)**: Siano date n variabili casuali indipendenti x_i , ciascuna con media μ_i e varianza σ_i^2 finite (non necessariamente uguali). Indipendentemente dalla forma particolare delle funzioni di distribuzione delle singole x_i , la somma $s = \sum_{i=1}^n x_i$ è asintoticamente normale nel limite $n \rightarrow \infty$, nel senso che per n grande la funzione di distribuzione di S tende ad una distribuzione di Gauss come media e varianza date da

$$E[S] = \sum_{i=1}^n \mu_i \quad e \quad Var(S) = \sum_{i=1}^n \sigma_i^2$$

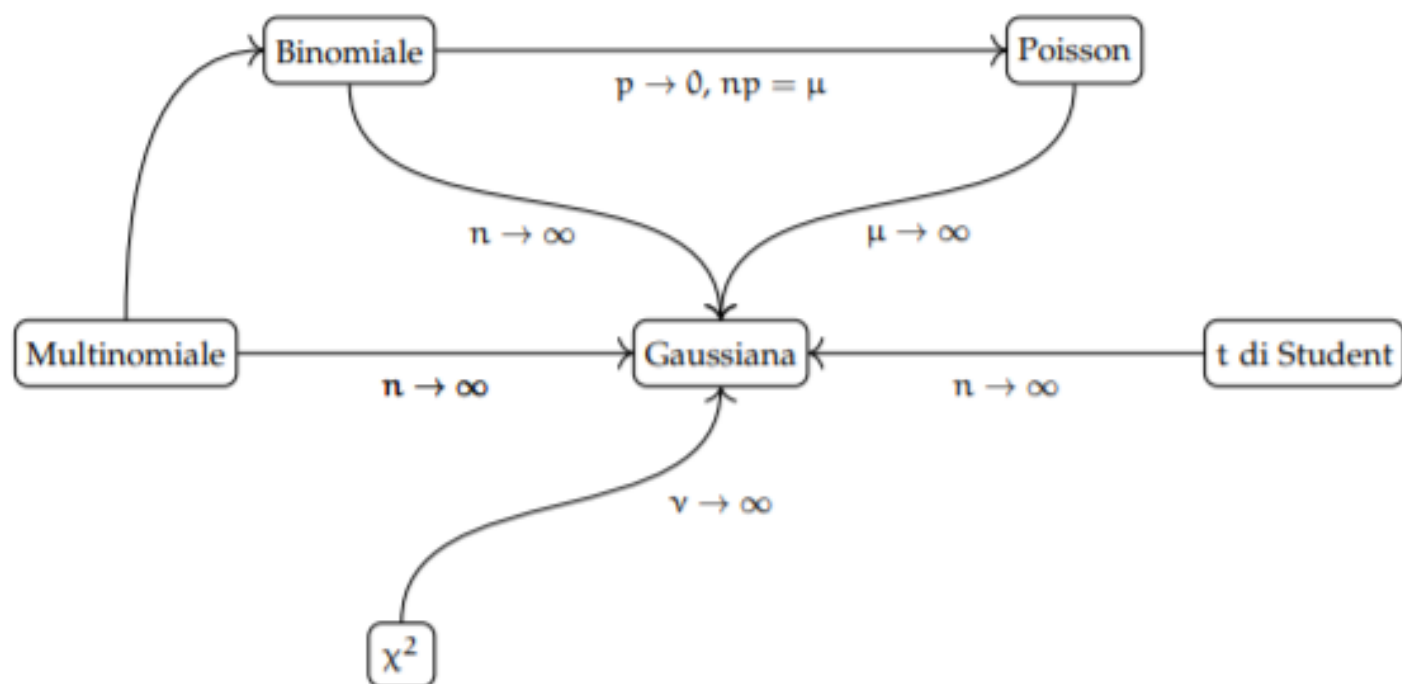
Cosa è allora la media?

Il teorema centrale del limite ci garantisce anche che la media (che, a parte un fattore costante di normalizzazione $1/n$ non è nient'altro che una somma) di un numero abbastanza grande di variabili casuali è distribuita approssimativamente come una Gaussiana. Questo implica che, quando scriviamo il risultato di una misura utilizzando la media e la deviazione standard della media, di fatto il livello di confidenza è fissato al 68

In breve

- In una situazione con due soli esiti possibili, la distribuzione binomiale descrive la probabilità di avere k successi in n tentativi, data la probabilità p di successo nell'evento elementare.
- La distribuzione di Poisson è, formalmente, il limite della binomiale per $p \rightarrow 0$ e $n \rightarrow \infty$, con $\mu = np$ costante. Essa descrive, cosa più importante, una larga classe di fenomeni, i fenomeni stazionari, e permette di rispondere alla domanda: "qual è la probabilità che si verifichino k eventi quando in media se ne verificano μ ?"
- La distribuzione costante è un buon modello per gli errori quando lo strumento di misura è digitale. (Tenete a mente il numero $\sqrt{12}$ perché lo ri-incontrerete nella vostra carriera di Fisici.)
- La distribuzione esponenziale, tra le altre cose, descrive la distribuzione delle differenze di tempi tra eventi successivi in un processo stazionario (cioè quando la statistica dei conteggi è Poissoniana). Essa ha anche l'interessante proprietà dell'assenza di memoria.
- La distribuzione di Gauss può essere ricavata come limite della distribuzione di Poisson per $\mu \rightarrow \infty$, ma è in realtà un attrattore (cioè un limite sotto condizioni opportune) per molte distribuzioni. L'integrale indefinito della Gaussiana non ha espressione analitica, per cui richiede di essere tabulato (vedi appendici A.1–A.5).
- La somma (e di conseguenza la media) di un numero abbastanza grande di variabili casuali tende, sotto ipotesi molto deboli, ad assumere una distribuzione Gaussiana indipendentemente dalla forma delle distribuzioni di partenza (teorema centrale del limite). Questo fissa il livello di confidenza (al 68%) quando scriviamo il risultato di una misura come $m \pm s_m$.

Distribuzione	Espressione	Media	Varianza	Skewness
Binomiale	$\mathcal{B}(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$
Poissoniana	$\mathcal{P}(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}$	μ	μ	$\frac{1}{\sqrt{\mu}}$
Uniforme	$u(x; a, b) = \begin{cases} \frac{1}{(b-a)} & a \leq x \leq b \\ 0 & x < a; x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	0
Esponenziale	$\varepsilon(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & 0 \leq x \leq \infty \\ 0 & x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2
Gaussiana	$N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2	0
χ^2	$p(x; \nu) = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} x^{\frac{\nu-2}{2}} e^{-\frac{x}{2}}$	ν	2ν	$\sqrt{\frac{8}{\nu}}$



Capitolo 6

Introduzione ai metodi di Monte Carlo

In questa unità si inizia parlando di numeri casuali, partendo dal metodo utilizzato da Archimede per approssimare π e spiegare come si fa adesso tramite numeri casuali ad approssimarlo, poi si guarda la macchina di Galton, ovvero una macchina utilizzata in laboratorio come una macchina dai risultati casuali. Una sequenza casuale è una sequenza non determinata e che non segue pattern ma è difficile darne una definizione poiché dato un campione di sequenze dove alcune possono seguire un pattern ed altre no ognuna ha la stessa probabilità di essere scelta, ma se viene data una sequenza che segue un pattern possiamo dire a priori che essa è casuale o non casuale? Un generatore di sequenza casuale può essere il rumore termico su una resistenza elettrica o il decadimento di un campione radioattivo. Un generatore di questo tipo si dice true random number generator (TRNG), mentre un generatore pseudo-casuale (PRNG) si basa su sequenze e pattern predeterminati che però data la grandezza non ne pregiudicano la casualità. Se dovessimo confrontare sommariamente le proprietà generali dei generatori casuali e pseudo-casuali diremmo essenzialmente tre cose:

- I generatori pseudo-casuali sono deterministici, mentre quelli casuali sono non deterministici. Nel primo caso la sequenza è univocamente determinata dal valore iniziale, e dunque prevedibile e riproducibile. In pratica la riproducibilità è una proprietà utile perché, come vedremo, ci permette di realizzare simulazioni in condizioni controllate.
- I generatori pseudo-casuali sono periodici, mentre quelli casuali sono aperiodici. La periodicità non è un problema in pratica a patto che il periodo sia abbastanza lungo. (Per completezza: il generatore pseudo-casuale della libreria standard di Python ha un periodo $2^{19937} - 1 > 10^{6000}$ e ci sono meno di 10^{27} nanosecondi nell'età dell'Universo, per cui la possibilità di esaurire la sequenza è abbastanza remota.)
- I generatori pseudo-casuali possono essere molto efficienti (in termini di numero di cifre prodotte nell'unità di tempo), mentre quelli casuali tendono ad essere comparativamente più lenti.

Sono infine discussi alcuni algoritmi pseudo casuali storici

Capitolo 7

Funzioni a variabili casuali

Questo capitolo riguarda lo studio delle probabilità data una variabile casuale come elemento di una funzione

Variabili casuali discrete

Se la funzione è biunivoca è facile pensare che

$$p(f(x)) = p(x)$$

come è altrettanto facile capire che il caso generale è invece

$$p(f(x)) = \sum p(x)$$

Variabili casuali continue

Essendo continue dobbiamo prendere un intervallo infinitesimale come ad esempio $[x - dx, x + dx]$ e se sono continue come per il caso discreto non ci sono problemi, dobbiamo solo effettuare un cambio di variabile da dx a dy (y avrà come intervallo $[dy-y, dy+y]$) dove

$$dy = \left| \frac{df}{dx}(x_0) \right| dx$$

quindi basterà imporre l'equazione

$$p_y(y)dy = p_x(x)dx = p_x(f^{-1}(y)) \left| \frac{df^{-1}}{dy}(x_0) \right| dy$$

se lasciamo l'ipotesi di biunivocità otteniamo

$$p_y(y) = \sum_{i=0}^m p_x(f^{-1}(y)) \left| \frac{df^{-1}}{dy}(x_0) \right| dy$$

distribuzione χ^2

La distribuzione χ^2 (chi-quadro) si indica solitamente con la seguente espressione

$$x = \sum_{i=1}^n z_i^2$$

con n che indica i gradi di libertà ed è la somma delle distribuzioni.

Come si scrive la funzione di distribuzione di x ? (Il problema non è proprio banale, ed abbiamo detto che la pertinenza con il filo conduttore di questo capitolo è limitata perché non si tratta di una funzione di una singola variabile casuale, ma della somma di variabili casuali diverse ed indipendenti.) Prima ancora di esaminare il problema in dettaglio, grazie al fatto che sappiamo tutto sulla funzione di distribuzione di ciascuna delle z_i^2 , possiamo dire un certo numero di cose su x .

$$E[x] = \sum_{i=1}^n E[z_i^2] = \sum 1 = n$$

$$Var(x) = \sum_{i=1}^n Var(x) = \sum 2 = 2n$$

Inoltre tramite il teorema del limite centrale sappiamo che la distribuzione tendereà ad una gaussiana per $n \rightarrow \infty$ Per competeza possiamo mostare senza dimostrazione l'espressione x^2

$$\varrho(x; n) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

La distribuzione di Cauchy

Data ϕ una variabile casuale continua nel intervallo $[-\pi/2, \pi/2]$

$$p_{\phi}(\phi) = \begin{cases} \frac{1}{\pi} & -\frac{\pi}{2} \leq \phi \leq \frac{\pi}{2} \\ 0 & \phi < -\frac{\pi}{2}; \phi > \frac{\pi}{2} \end{cases}$$

siamo interessati a calcola di probabilità del coefficente angolare, ovvero

$$x = f(t) = \gamma \tan \phi + x_0$$

La forma sarà

$$c(x; x_0, \gamma) = \frac{1}{\pi} \left[\frac{\gamma}{(x - x_0)^2 + \gamma^2} \right]$$

dove se $\gamma = 1$ e $x_0 = 0$ è detta in forma normale.

Capitolo 8

Metodi di fit

Descrizione del problema generale

Dato un insieme di dati sperimentali, il problema di fit o best fit consiste nel cercare una funzione (o modello), possibilmente semplice, che descriva in modo soddisfacente i dati stessi. Esistono 2 problemi a riguardo: il primo è, appunto, la ricerca del modello specifico che meglio si adatta ai nostri dati secondo una qualche metrica ragionevole. Il secondo è valutare a posteriori il livello di accordo (o la bontà del fit) per capire se il nostro modello specifico è un buon modello oppure no. Più praticamente quello che dobbiamo fare è: data una serie di n coppie ordinate di misure di due generiche grandezze x ed y , con le rispettive incertezze

$$(x_i \pm \sigma_{x_i}, y_i \pm \sigma_{y_i}) \quad i = 1 \dots n$$

ed una famiglia di funzioni, dipendente da un numero m di parametri $\theta_1 \dots \theta_m$, che pensiamo possa rappresentare la dipendenza di y da x

$$y = f(x; \theta_1 \dots \theta_m)$$

il nostro primo compito è dunque quello di trovare i valori $\tilde{\theta}_1 \dots \tilde{\theta}_m$ dei parametri in modo da massimizzare (in un senso che preciseremo tra un attimo) l'accordo tra modello e dati. La famiglia di funzioni scelta va presa in base alla teoria fisica delle osservazioni fatte, infatti dati n punti sul grafico esiste sempre una funzione di grado $n-1$ passante per tutti i punti, tuttavia questa funzione oltre a levare le incertezze di misura non aggiunge informazioni restituendo lo stesso numero di valori inseriti.

Introduzione ai fit dei minimi quadrati

Facciamo 3 ipotesi:

1. le n misure sono tra loro indipendenti;
2. i valori y_i sono misurati in corrispondenza di x_i noti, ovverosia le incertezze di misura sugli x_i sono trascurabili;
3. gli errori di misura sulla variabile dipendente sono Gaussiani, cioè gli y_i tendono a fluttuare attorno al misurando con una distribuzione Gaussiana con deviazione standard σ_{y_i} nota a priori.

Il secondo punto dice semplicemente che la l'errore di x è molto inferiore della propagazione dell'errore x su y , in formule ricordandoci che la formula generale della propagazione dell'errore nel caso unidimensionale è lo sviluppo di Taylor fermato al primo ordine $f(x \pm \Delta x) = f(x) + \frac{df}{dx}(x)\Delta x$

$$\left| \frac{df}{dx} x_i; \tilde{\theta}_1 \dots \tilde{\theta}_m \right| \sigma_{x_i} \ll \sigma_{y_i} \quad i = 1 \dots n$$

Costruiamo adesso una somma χ^2

$$\chi^2(\theta_1 \dots \theta_m) = \sum_{i=1}^n \left(\frac{y_i - f(x_i; \theta_1 \dots \theta_m)}{\sigma_{y_i}} \right)^2$$

A questo livello il χ^2 non è altro che la somma dei quadrati delle differenze tra dati e modello, misurate in unità di barre d'errore (come per la varianza il quadrato serve per far sì che fluttuazioni positive e negative si eliminino a vicenda). È chiaro allora che questa quantità misura, in un qualche senso, la distanza complessiva tra una serie di dati ed un modello, in funzione del valore dei parametri da cui il modello dipende: più piccolo è il χ^2 , migliore l'accordo tra dati e modello. Ora

che abbiamo una misura oggettiva dell'accordo tra dati e modello, possiamo trovare la funzione di best fit semplicemente minimizzando il nostro χ^2 rispetto agli m parametri del modello

$$\begin{cases} \frac{\partial \chi^2}{\partial \theta_1}(\tilde{\theta}_1 \dots \tilde{\theta}_m) = 0 \\ \vdots \\ \frac{\partial \chi^2}{\partial \theta_m}(\tilde{\theta}_1 \dots \tilde{\theta}_m) = 0 \end{cases}$$

questo sistema fornisce un insieme di m equazioni che, almeno in linea di principio, ci permettono di ricavare gli m valori di best fit $\tilde{\theta}_1 \dots \tilde{\theta}_m$ per i parametri.

Fit dei minimi quadrati nel caso costante: la media pesata

L'esempio più facile di fit ad un parametro (q) è quello con una funzione costante, cioè indipendente dalla variabile

$$f(x; q) = q$$

riproducendo i passaggi generali detti prima otteniamo $\chi^2(q) = \sum \left(\frac{y_i - q}{\sigma_{y_i}} \right)^2 \rightarrow \frac{\partial \chi^2}{\partial q}(\tilde{q}) = -2 \sum \left(\frac{y_i - \tilde{q}}{\sigma_{y_i}^2} \right) = 0 \rightarrow \tilde{q} = \frac{\sum \frac{y_i}{\sigma_{y_i}^2}}{\sum \frac{1}{\sigma_{y_i}^2}}$
sostituendo $\frac{1}{\sigma_{y_i}^2}$ con w_i si capisce l'espressione "media pesata"

$$\tilde{q} = \frac{\sum w_i y_i}{\sum w_i}$$

è bene sapere che i pesi statisticamente corretti da assegnare alle misure sono dunque gli inversi dei quadrati delle rispettive incertezze, il che, almeno a livello qualitativo, ha senso in quanto le misure con le incertezze più grandi debbono contare meno nella media. calcoliamo ora l'errore nella nostra stima \tilde{q} di q

$$\sigma_q^2 = Var\left(\frac{\sum w_i y_i}{\sum w_i}\right) = \frac{1}{\sum \frac{1}{\sigma_{y_i}^2}}$$

σ_q è strettamente minore di ciascun σ_{y_i} come è lecito aspettarsi sulla base del fatto che tutte le misure, più o meno a seconda dell'incertezza associata, contribuiscono informa.

Caso lineare

Un altro esempio, questa volta con più di un parametro (q, m). Salterò i passaggi algebrici scrivendo solo le equazioni iniziali ed i risultati

$$f(x; mq) = mx + q \quad \chi^2(q, m) = \sum_{i=1}^n \left(\frac{y_i - mx_i - q}{\sigma_{y_i}} \right)^2$$

$$\begin{cases} \frac{\partial \chi^2}{\partial q}(\tilde{q}, \tilde{m}) = -2 \sum_{i=1}^n \left(\frac{y_i - \tilde{m}x_i - \tilde{q}}{\sigma_{y_i}^2} \right) = 0 \\ \frac{\partial \chi^2}{\partial m}(\tilde{q}, \tilde{m}) = -2 \sum_{i=1}^n \left(\frac{y_i - \tilde{m}x_i - \tilde{q}}{\sigma_{y_i}^2} \right) x_i = 0 \end{cases}$$

$$\tilde{q} = \frac{S_{xy}^0 S_x^2 - S_{xy}^1 S_x^1}{D} \quad \sigma_q^2 = \frac{S_x^2}{D}$$

$$\tilde{q} = \frac{S_{xy}^1 S_x^0 - S_{xy}^0 S_x^1}{D} \quad \sigma_m^2 = \frac{S_x^0}{D}$$

risolto con le matrici e con $S_x^0 = \sum \frac{1}{\sigma_{y_i}^2}$, $S_x^1 = \sum \frac{x_i}{\sigma_{y_i}^2}$, $S_x^2 = \sum \frac{x_i^2}{\sigma_{y_i}^2}$, $S_{xy}^0 = \sum \frac{y_i}{\sigma_{y_i}^2}$, $S_{xy}^1 = \sum \frac{x_i y_i}{\sigma_{y_i}^2}$ e $D = S_x^0 S_x^2 - (S_x^1)^2$.

Approfondimenti sui fit dei minimi quadrati

- Si osserva che in un fit lineare c'è correlazione tra m e q .
- Per i fit con un polinomio arbitrario si procede con il metodo generale e possiamo creare una matrice come nel caso lineare per semplificare i calcoli.
- se σ_{y_i} non dipende dall'indice (cioè hanno lo stesso valore σ_y per tutti i punti sperimentali) allora $\sum (y_i - f(x; \theta_1 \dots \theta_m))^2$ e si chiama caso non pesato. Ad esempio, come logico che sia, il caso ad un parametro diventa la media aritmetica $\tilde{q} = \frac{1}{n} \sum y_i$ e $\sigma_q = \frac{\sigma_y}{\sqrt{n}}$

Il test del χ^2

Adesso che abbiamo inquadrato il metodo con cui si individuano i fit (che per la stramaggioranza delle volte non servirà poiché ci pensa python ad effettuare il tutto) possiamo guardare la seconda domanda, come si capisce su un fit è soddisfacente oppure no? Se il processo per minimizzare χ^2 ci forniva e i valori ottimali dei parametri, la risposta a questa altra domanda deve essere in qualche modo contenuta nel valore numerico del χ^2 stesso che il fit restituisce in corrispondenza del minimo

$$\chi^2(\theta_1 \dots \theta_m) = \sum_{i=1}^n \left(\frac{y_i - f(x_i; \theta_1 \dots \theta_m)}{\sigma_{y_i}} \right)^2$$

Il contenuto fisico del χ^2 : quanti gradi di libertà?

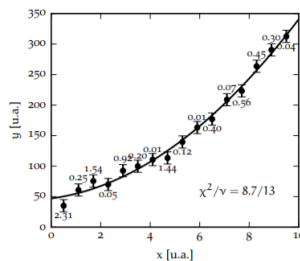
Immaginiamo di avere $\overline{\theta_1 \dots \theta_m}$ misurati perfettamente allora

$$\zeta_i = \frac{y_i - f(x_i; \overline{\theta_1 \dots \theta_m})}{\sigma_{y_i}}$$

sarà distribuito per definizione come una variabile Gaussiana in forma standard. Ora, dato che ogni ζ_i è elevato al quadrato, la nostra somma è, formalmente, una somma di n variabili Gaussiane in forma standard al quadrato, e noi abbiamo calcolato la sua funzione di distribuzione: una distribuzione del χ^2 ad n gradi di libertà il cui valore di aspettazione è n e la varianza è $2n$. Possiamo chiederci se tutto questo valga anche se calcoliamo la nostra somma in corrispondenza dei valori di best fit dei parametri anziché nei valori (che peraltro non conosciamo) veri. Ebbene, sotto ipotesi relativamente deboli la quantità $\chi^2(\theta_1 \dots \theta_m)$ è ancora distribuita come un χ^2 (cosa che non dimostreremo), ma il numero rilevante di gradi di libertà non è, in generale, pari al numero n di misure. Immaginiamo adesso un caso reale, essendo $y=q$ un caso particolare di $y=mx+q$ allora ovviamente sarà più accurato ed il minimo di χ^2 del fit lineare sarà minore o uguale di quello costante, un fit con $n-1$ gradi passerà per tutti i punti quindi $\chi^2=0$. Da un punto di vista dei principi primi, il numero di gradi di libertà del nostro problema è, per definizione, pari al numero di termini indipendenti nella somma, allora è chiaro che per ogni parametro che fittiamo dai nostri dati introduciamo un vincolo, cioè una relazione tra i dati stessi che porta alla diminuzione di un'unità nel numero di gradi di libertà. In definitiva il numero di gradi di libertà è pari al numero dei punti sperimentali n meno il numero di parametri m che stimiamo dal fit

$$\nu = n - m$$

Test del χ^2



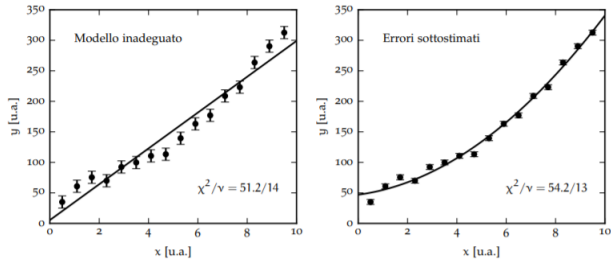
In figura è mostrato un fit ad una serie di 16 punti sperimentali con un modello quadratico $f(x) = lx^2 + mx + q$. Il modello ha 3 parametri per cui sappiamo che il numero di gradi di libertà del problema è $\nu = 16 - 3 = 13$. Ad occhio il fit sembra accurato: i punti sembrano oscillare attorno al modello di best fit con differenze dell'ordine delle barre d'errore; 7 punti stanno sotto al modello e 9 sopra, dove in media ce ne aspetteremmo 8 e 8, e non si notano regioni in cui il modello

sovrastima o sottostima sistematicamente i dati; vi sono esattamente 3 punti che distano più di una barra d'errore dal modello, quando in media ce ne aspettiamo il 32%, cioè circa 5. Vediamo adesso che il valore χ^2 restituito dal fit è 8.7, essendo i gradi di libertà 13 la media è 13 e la deviazione è $\sqrt{23} = 5.10$ e di conseguenza ci troviamo entro una deviazione standard, inoltre le tabelle del χ^2 dicono che

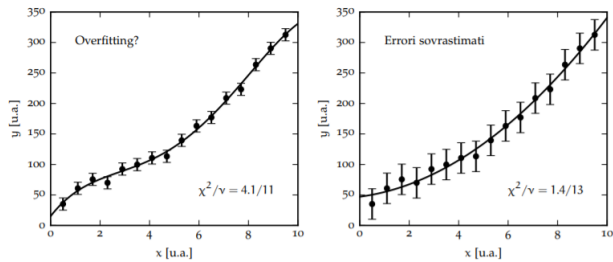
$$P(\chi^2_{13} \leq 8.7) \approx 20\%$$

ovvero se ripetessimo molte volte l'esperimento nel 20% dei casi otterremo un χ^2 minore e nell'80% dei casi un χ^2 maggiore, ma il punto è che nessuno di questi due numeri è troppo vicino a 0, per cui possiamo concludere che il nostro fit è un buon fit (come dedotto "ad occhio"); dire che un fit è un buon fit equivale a dire che non abbiamo ragioni per rigettare il nostro modello, ma non abbiamo mai la certezza di aver fittato correttamente.

Osserviamo ora i casi in cui non va bene:



nel primo caso ci aspettiamo $\chi^2 = 14$ con deviazione 5.29 mentre nel secondo 13 con deviazione 5.10, è palese come questo sia invece a 7-8 deviazioni in entrambi i casi e seguendo le tabelle la in entrambi i casi $P > 99.99\%$ e quindi la possibilità che ripetendo le misure queste abbiano un χ^2 più basso è del 99.99%. A questo punto o siamo stati incredibilmente sfortunati, oppure una delle nostre due assunzioni (o entrambe) non è verificata. Da Fisici, optiamo per la seconda, se il χ^2 è troppo grande si hanno due casi, che sono legati uno ad uno alle nostre due ipotesi di lavoro: o il numeratore della somma è troppo grande (cioè il nostro modello non è adeguato a fittare i dati) oppure il denominatore è troppo piccolo (cioè gli errori sono sottostimati). E se invece χ^2 è troppo piccolo?

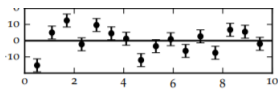


Nella prima figura il modello è un polinomio di 4 grado e dalle tabelle si ottiene $P(\chi^2_{11} \leq 4.1) \approx 2\%$ e $P(\chi^2_{13} \leq 1.4) \approx 0.5\%$. Cioè: se ripetessimo l'esperimento e se il modello fosse corretto e gli errori stimati correttamente, la probabilità di ottenere un χ^2 pari a quello ottenuto o più piccolo sarebbe molto piccola e otterremmo quasi sempre un χ^2 più grande, questo può essere dato o dagli errori sovrastimati oppure dal modello troppo flessibile (overfitting) che è il motivo per cui abbiamo detto all'inizio che un modello di grado $n-1$ passante per tutti i punti sarebbe generalmente scorretto poiché a meno che i dati non siano perfetti ripetendo le misure otterremmo un fitto "cattivo".

Il grafico dei residui

È possibile fare un fit dove mettiamo il modello come una costante e i dati con i relativi errori a definirla come una media pesata, in questo modo è più facile osservare la bontà del fit.

$$r_i = y_i - f(x_i; \tilde{\theta}_1 \dots \tilde{\theta}_m) \quad e \quad \sigma r_i = \sigma y_i$$



Fit dei minimi quadrati e test del χ^2 per una distribuzione

Supponiamo adesso di voler fare un fit dei minimi quadrati o un test del χ^2 nel caso di una distribuzione, ad esempio, potremmo lanciare un dado a sei facce per 100 volte e vedere se l'ipotesi che il dado sia equo può essere rigettata. Oppure potremmo voler confrontare in senso statistico un istogramma con una distribuzione di variabile continua. È chiaro che questo problema abbia a che fare con quanto detto fin ora, ma la differenza fondamentale è che qui non abbiamo a che vedere con misure sperimentali, ma con conteggi. Cominciamo con l'indicare con o_i le occorrenze osservate in corrispondenza dell' i -esimo valore della nostra variabile discreta o dell' i -esimo canale del nostro istogramma. Il modello ci permetterà in generale di calcolare le occorrenze attese e_i nella forma

$$e_i(\theta_1 \dots \theta_m) = \begin{cases} NP(x_i; \theta_1 \dots \theta_m) \\ N \int_{x_i}^{x_{i+1}} p(x; \theta_1 \dots \theta_m) dx \end{cases}$$

dove l'ultimo integrale è inteso tra gli estremi dell' i -esimo canale dell'istogramma ed N è il numero totale $N = \sum_i o_i$ o di occorrenze osservate. (Nel caso dei nostri 100 lanci di un dado le occorrenze osservate o_i sono banalmente il numero di volte in cui ogni faccia è uscita. Se ipotizziamo che il dado sia equo, il nostro modello prevede che le occorrenze attese siano identiche tra loro e pari a $e_i = 100/6$.) Con questi ingredienti possiamo scrivere quella che anticipiamo essere la risposta corretta alla nostra domanda iniziale, ovvero:

$$\chi^2 = \sum_{i=1}^n \frac{(o_i - e_i)^2}{e_i}$$

Quest'ultima equazione ricorda molto quella classica del χ^2 , infatti $o_i \rightarrow y_i$, $e_i(\theta_1 \dots \theta_m) \rightarrow f(x_i; \theta_1 \dots \theta_m)$ ovvero immagine e previsione del modello, infine al denominatore abbiamo bisogno di una stima dell'errore, ma infatti $\sigma_{e_i} = \sqrt{e_i}$ e $\sigma_{e_i}^2 = e_i$. Nei casi in cui il numero totale N di eventi (o conteggi) è fissato gli n elementi della somma non sono più tutti indipendenti tra di loro, perché non sono più indipendenti tra di loro le grandezze o_i . Se conosciamo il valore di $n-1$ occorrenze possiamo calcolare banalmente la rimanente attraverso

$$o_k = N - \sum_{i \neq k} o_i \quad \text{ad esempio} \quad o_n = N - \sum_{i=0}^{n-1} o_i$$

inoltre per le distribuzioni

$$\nu = n - m - 1$$

Ovvero: il numero di gradi di libertà è pari al numero dei dati sperimentali n meno il numero di parametri m che stimiamo dal fit meno 1.

Poissoniana o binomiale? Un semplice esempio

Osservando l'analogia tra l'equazione del χ^2 classica e quella per le distribuzioni si intuisce facilmente che si tratta di una distribuzione poissoniana, se tuttavia il numero di eventi N è conosciuto allora si tratta di una binomiale poiché diventa uno contro tutti (ripetizione o non ripetizione del precedente evento). Ad esempio lanciamo un dado a 6 facce per 100 volte, il numero di volte o_2 in cui esce la faccia 2 sarà una variabile casuale distribuita secondo una binomiale $B(o_2; 100, 1/6)$. Il valore di aspettazione e la varianza sono

$$E[o_i] = Np_i = e_i \quad \text{e} \quad \text{Var}(o_i) = Np_i(1 - p_i) \neq e_i$$

Il problema sta proprio nel fatto che in questo caso la varianza delle occorrenze attese è diversa dal valore di aspettazione delle occorrenze stesse, tuttavia seguendo un calcolo che non riporto qua (p.178 statnotes baldini) $(1 - p_i)$ viene esplicitato.

Il concetto di p-value

Ricordiamo cosa abbiamo fatto in questa sezione fin ora: quando facciamo il test χ^2 di verifica delle ipotesi ci aspettiamo un valore uguale ai gradi di libertà ν con una deviazione $\sqrt{2\nu}$, se il risultato è entro $2/3$ deviazioni dalla media non abbiamo ragione di rigettare il modello. Se χ^2 è troppo grande vuol dire che gli errori sono sottostimati oppure il modello è incorretto, se è troppo piccolo ci sono troppi parametri o gli errori sono sovrastimati. Il p-value è la probabilità di ottenere un risultato più estremo (1-p se il χ^2 è più grande del 50%, altrimenti è p). A causa delle varie interpretazioni che gli sono state date tuttavia, il p-value è bandito dal mondo della statistica poiché spesso la sua malinterpretazione aveva motivi fraudolenti, in particolare questo non è:

- Il p-value non è la probabilità che l'ipotesi nulla sia corretta (dove l'ipotesi nulla H_0 è ad esempio l'equità di un dado o di una moneta).
- il p-value non è la probabilità che l'ipotesi alternativa ad H_0 sia errata
- Il p-value non è la probabilità che quello che abbiamo osservato sia semplicemente l'effetto di una fluttuazione statistica.
- Il p-value non è la probabilità di rigettare per sbaglio l'ipotesi nulla.

Fit dei minimi quadrati e test del χ^2 con errori non Gaussiani

Nonostante abbiamo detto che gli errori debbano essere gaussiani, questo è un punto che spesso viene tralasciato e spesso vengo utilizzate incertezze indipendentemente dal loro tipo. Ovviamente il problema può essere facilmente ovviato utilizzando il metodo di monte carlo per portare l'incertezza ad una gaussiana ma spesso non importa e possiamo procedere al χ^2 con un qualsiasi errore. Innanzitutto l'equazione del chi quadro rimane una misura ragionevole della distanza complessiva tra dati e modello nella maggior parte dei casi, per cui i valori dei parametri che si ottengono dalla sua minimizzazione sono per lo più ragionevoli anche se le incertezze di misura non sono Gaussiane (fanno eccezione i casi in cui gli errori di misura sono notevolmente asimmetrici, ciò si presenta quando in una distribuzione ha pochi conteggi). Il calcolo delle incertezze di misura per i modelli di fit si basa essenzialmente sul fatto che per variabili casuali indipendenti la varianza della somma è uguale alla somma delle varianze, il che, è vero indifferentemente dalla distribuzione generatrice. Il test del χ^2 è il punto in cui sorgono i veri problemi, infatti se gli errori non sono gaussiani allora per definizione l'equazione del χ^2 non è distribuita come tale. Se il numero di gradi di libertà del problema è abbastanza grande, il teorema del limite centrale ci viene di nuovo in aiuto garantendo che la distribuzione della nostra somma sia asintoticamente Gaussiana, ma la sua varianza non sarà più, in generale, 2ν .

Di fatto noi sappiamo la risposta esatta almeno per un caso praticamente rilevante, ovvero quello di errori uniformi (per esempio se utilizziamo uno strumento digitale ed assumiamo come incertezza la risoluzione strumentale diviso $\sqrt{12}$). Nella sezione 7.3.1 abbiamo visto che la varianza del quadrato di una distribuzione uniforme con varianza 1 è $4/5$, per cui in questa situazione avremo

$$Var(\chi^2(\tilde{\theta}_1 \dots \tilde{\theta}_m)) = \frac{4}{5}\nu$$

Meno della metà della varianza della corrispondente distribuzione del χ^2 : se convertiamo in probabilità il valore del χ^2 ottenuto con errori non Gaussiani possiamo sbagliare di grosso. In questo caso se ν è abbastanza grande possiamo utilizzare l'approssimazione Gaussiana (con la deviazione standard opportuna). Nel caso generale è necessario calcolare la densità di probabilità rilevante: analiticamente oppure attraverso una simulazione Monte Carlo.

Il fit in pratica: metodi numerici

Prendiamo la famiglia di funzioni

$$f(x; \omega) = \sin(\omega x)$$

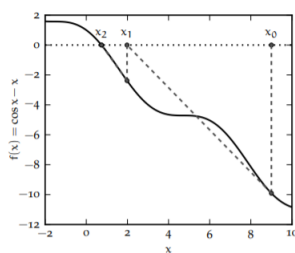
minimizzando il χ^2 di tale funzione osserviamo che esso non è risolvibile analiticamente in forma chiusa (Una soluzione in forma chiusa è qualsiasi valore/formula esatto e non approssimato che può essere valutata in un numero finito di operazioni standard). La differenza tra questa e le precedenti equazioni del χ^2 è che questa è non lineare nei parametri e si parla in questo caso di fit dei minimi quadrati non lineare; la tecnica standard per affrontare il problema è quello dell'utilizzo di metodi numerici iterativi.

Digressione: il metodo di Newton in una dimensione

Immaginiamo di voler risolvere un'equazione con $x-x=0$ possiamo utilizzare il metodo delle tangenti (la funzione deve essere uguale a 0), simile a quello della bisezione ma più veloce, unica controindicazione è che può avere un massimo di iterazioni data la sua formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

infatti se la derivata dovesse diventare 0 allora x divergerebbe (prende spunto dalla serie di Taylor bloccata al primo ordine ($f(x) = x_0 - \frac{f(x_0)}{f'(x_0)}x$)). Visualmente si ottiene il seguente schema.



Minimi quadrati non lineari e metodi iterativi

Quando lavoriamo su python questo non utilizza i metodi utilizzati per il caso lineare all'inizio del capitolo ma utilizza piuttosto metodi iterativi come il metodo di bisezione o quello delle tangenti. Noi utilizzeremo la libreria *scipy.optimize* attraverso la funzione *curve_fit*

$popt, pcov = \text{scipy.optimize.curve_fit}(f, xdata, ydata, p0 = \text{None}, \text{sigma} = \text{None})$

f=la funzione.

xdata= il vettore delle coordinate x_i delle misure.

ydata= il vettore delle coordinate y_i delle misure.

p0= i valori iniziali dei parametri del modello, come ad esempio il seme della successione del metodo delle tangenti. (non obbligatorio)

sigma= le incertezze di misura σ_{y_i} sui valori y_i .

popt= un array di numpy di lunghezza *m* contenente i valori ottimali dei parametri del modello. In altre parole queste sono proprio le stime dei parametri che cerchiamo.

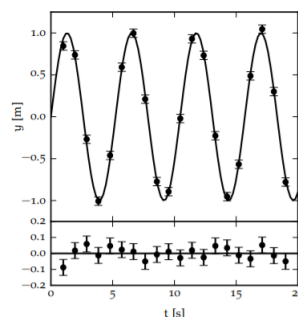
pcov= la matrice di covarianza $m \times m$ tra i parametri del modello. Ricordiamo che le incertezze sui parametri sono le radici quadrate degli elementi diagonali della matrice stessa.

Es. con il seguente script otteniamo il seguente plot.

```

1  https://bitbucket.org/.../fit_sin.py
2  import numpy
3  from scipy.optimize import curve_fit
4
5  def f(x, omega):
6      """Definition of the fit function.
7
8      Note that x goes first, then come the parameters.
9
10     """
11     return numpy.sin(omega*x)
12
13 # Definition of the input data set.
14 numpy.random.seed(100)
15 n = 20
16 dy = 0.05
17 omega = 1.2
18 x = numpy.linspace(1., 19., n)
19 y = f(x, omega) + numpy.random.normal(0, dy, size=n)
20 # Perform the fit and calculate the chisquare
21 popt, pcov = curve_fit(f, x, y, 1., dy)
22 chisq = ((y - f(x, popt))/dy)**2).sum()
23 # Print the fit output.
24 print('omega = %.4f +/- %.4f' %
25       (popt, numpy.sqrt(pcov.diagonal())))
26 print('Chisquare = %.1f' % chisq)
27 [Output]

```



Il problema dei valori iniziali

Quando si parlava di minimo del χ^2 si prende l'equazione in funzione dei parametri e dopo averne fatto la derivata si pone uguale a 0 (ciò significa che se abbiamo una media pesata ci troviamo di fronte ad una parabola). Il problema sorge quando abbiamo modelli più complicati e abbiamo più minimi, infatti i metodi iterativi ci faranno convergere al minimo più vicino che non necessariamente è quello assoluto (quello che noi cerchiamo) e fanno capire l'importanza del valore iniziale. Per capire se il minimo trovato non è quello corretto usiamo il χ^2 ed il grafico dei residui e poi cerchiamo un valore iniziale più corretto (trovandolo guardando i dati ed il modello). Un esempio di fit complesso su cui sviluppare i ragionamenti detti è portato nelle dispense del Baldini a pagg. 187.

Fit di tipo generale

Cosa succede se l'errore propagato sulla *x* non è trascurabile? Ovviamente ciò che abbiamo detto sul χ^2 non è più valido e bisogna trovare un altro metodo. Tramite python ed il pacchetto di *scipy* *odr* (orthogonal distance regression) possiamo

calcolare l'errore come una diagonale del rettangolo di lati x , y . Un metodo analitico sta nel modificare il χ^2 adeguatamente ma non lo trascriverò (stat notes p. 192).

Capitolo 9

Cenni al principio di massima verosimiglianza

La verosimiglianza (maximum liklyhood) è la risposta all'quesito che ci siamo posti nel capitolo 3 ovvero la probabilità che dato dato un numero di eventi conosciuti si riescano a recuperare i dati iniziali.

Stime di massima verosimiglianza

Cerchiamo di formalizzare il problema in termini più generali. Sia dato un insieme di n misure sperimentali $x_1...x_n$ ed un modello dipendente da m parametri $\theta_1...\theta_m$. Fissato un generico vettore di valori dei parametri il modello definisce univocamente la probabilità (o la densità di probabilità, nel caso i dati siano variabili continue) di ottenere proprio le nostre misure. Adesso noi scriviamo la funzione di verosimiglianza come

$$L(\theta_1...\theta_m; x_1...x_n) = p(\theta_1...\theta_m | y_1...y_n)$$

Il principio di massima verosimiglianza si può enunciare dicendo che la miglior stima $\tilde{\theta}$ del vettore dei parametri θ è quella che massimizza la funzione di verosimiglianza:

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial \theta_1}(\tilde{\theta}_1...\tilde{\theta}_m; x_1...x_n) = 0 \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial L}{\partial \theta_m}(\tilde{\theta}_1...\tilde{\theta}_m; x_1...x_n) = 0 \end{array} \right.$$

espliciterò adesso le formule di massima verosimiglianza per le varie distribuzioni senza spiegare il ragionamento che c'è dietro (ad ogni modo si trovano da p. 198).

Stime di massima verosimiglianza nel caso poissoniano

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n k_i$$

Stime di massima verosimiglianza nel caso esponenziale

$$\tilde{\mu} = \frac{1}{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$$

Stime di massima verosimiglianza nel caso Gaussiano

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$
$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{\mu})^2$$

Stime di massima verosimiglianza nel caso uniforme

$$\tilde{\mu} = (\min_i x_i + \max_i x_i)$$

Incertezze nelle stime di massima verosimiglianza

Formula del caso unidimensionale

$$L(\theta) = \frac{1}{\sigma_{\tilde{\theta}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\theta - \tilde{\theta}}{\sigma_{\tilde{\theta}}}\right)^2}$$

Dove $\sigma_{\tilde{\theta}}$ è proprio l'incertezza sulla stima $\tilde{\theta}$ di θ cui siamo interessati.

Il χ^2 come applicazione del principio di massima verosimiglianza

Si osserva che per il χ^2 la massima verosimiglianza è uguale alla minimizzazione del χ^2 per costruire i fit.

Ottica

pietro.faraggiana

May 2021

Indice di rifrazione

Sappiamo che la luce si propaga sottoforma di onda elettromagnetica, definiamo indice di rifrazione del mezzo n la quantità:

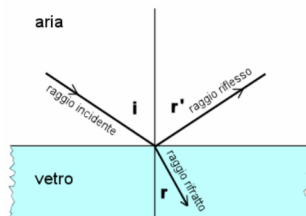
$$v = \frac{c}{n}$$

dove v è la velocità della luce nel vuoto c è $3 \cdot 10^8 m/s$ e n è l'indice di rifrazione (1 nel vuoto, 1.0003 nell'aria, 1.33 nell'acqua...). La legge empirica di Gladstone lega l'indice di rifrazione alla densità (e poi alla pressione nei gas tramite l'equazione di stato dei gas ideali)

$$n - 1 = k\rho \rightarrow \frac{n - 1}{n_0 - 1} = \frac{P}{P_0} \frac{T_0}{T}$$

Riflessione e rifrazione

Nelle più semplici applicazioni dell'ottica le grandezze in gioco sono molto maggiori della lunghezza d'onda della radiazione considerata, e quindi possiamo assumere che la luce si propaghi in linea retta, nonostante il suo comportamento ondulatorio. Parliamo perciò di ottica geometrica. Quando un raggio incontra una superficie trasparente esso viene rifratto (a meno che non sia parallelo alla normale $i=0$) e riflesso con $i = i'$ angolo tra il raggio e la normale.



La legge di Snell fornisce la relazione fra gli angoli di incidenza i e di rifrazione r :

$$n \sin i = n' \sin r$$

dove n e n' sono gli indici di rifrazione. Dalla formula di Snell 'e possibile ricavare l'angolo critico i_c oltre il quale si avrà riflessione totale. Questo angolo è tale per cui $r = \pi/2$, per cui:

$$i_c = \arcsin\left(\frac{n'}{n}\right)$$

Per $i > i_c$ si ha riflessione totale, ma questo può avvenire solo se $n' < n$, cioè nel passaggio da un mezzo più denso a uno meno denso.

Cammino ottico

Il cammino d di un raggio di luce che si propaga in un mezzo con indice di rifrazione n , vale:

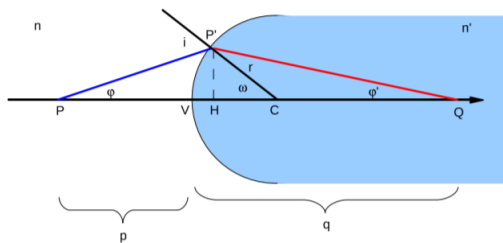
$$d = \frac{c}{n}t \rightarrow l = nd = ct$$

l è detto cammino ottico e rappresenta lo spazio percorso dalla luce nel vuoto mentre percorre un cammino lungo vari mezzi di diverso indice di rifrazione. Nel 1650 Fermat annuncia il seguente principio

Il percorso fra due punti preso da un raggio di luce 'e quello che 'e attraversato nel minor tempo.

derivando il cammino ottico e cercando il minimo si trova la legge di Snell-Cartesio, l'invarianza nell'invertire il cammino e $i = i'$

Diottro sferico



Si indica con il termine diottro sferico una calotta sferica che separa due mezzi con indice di rifrazione diverso. L'asse che congiunge il vertice V al centro di curvatura C è chiamato asse principale del diottro. P è il punto-oggetto, Q il punto-immagine e sono i punti coniugati. L'equazione che descrive i raggi che interagiscono con il diottro sferico è la seguente

$$\frac{n}{p} + \frac{n'}{q} = \frac{n' - n}{r}$$

che è nota con il nome di formula di Gauss per il diottro sferico. L'approssimazione applicata è nota come approssimazione di Gauss. Essa consiste non

soltanto nel considerare raggi che formano un angolo piccolo con l'asse ottico, ma che oltre a ciò incidono in un'area ristretta intorno al vertice della calotta sferica. Perciò dato un punto-oggetto che si trovi fuori dall'asse ottico, ma non troppo distante da esso, la formula di Gauss per il diottro sferico può servire a calcolare la distanza del punto-immagine. In definitiva possiamo ragionevolmente assumere che solamente in approssimazione di Gauss tutti i raggi uscenti da P vanno ad incontrarsi in Q.

Definizione dei fuochi

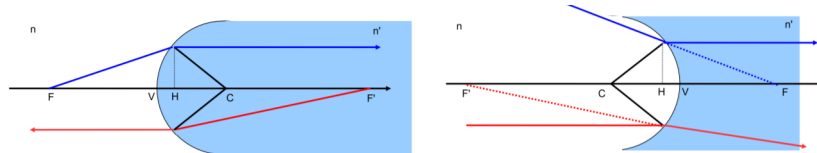
Ipotizziamo che $q \rightarrow \infty$ si ottiene

$$\frac{n}{f} = \frac{n' - n}{r}$$

dove $f=p$. Se invece $p \rightarrow \infty$

$$\frac{n'}{f'} = \frac{n' - n}{r}$$

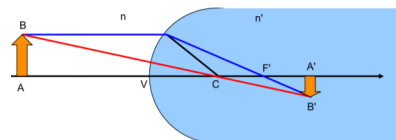
dove $f'=q$. Le quantità f e f' vengono chiamate fuochi del diottro



da cui $\frac{f'}{f} = \frac{n'}{n}$ cioè la distanza del fuoco dal vertice del diottro è tanto maggiore quanto maggiore è il valore dell'indice di rifrazione del mezzo. Le posizioni dei fuochi sono simmetriche solo nel caso in cui i due mezzi abbiano lo stesso indice di rifrazione, cosa che si verifica nel caso in cui la superficie sferica del diottro abbia uno spessore molto sottile rispetto alle distanze in gioco.

Formazione dell'immagine

Per costruire l'immagine di un oggetto esteso possiamo sfruttare le proprietà appena viste del diottro. Assumiamo che un oggetto AB si trovi su un piano ortogonale all'asse ottico e abbia l'estremo A sopra di esso. Tracciamo una retta da B al diottro parallela all'asse e poi al fuoco e da B al centro del diottro, il punto in cui si incontrano è l'immagine di B.



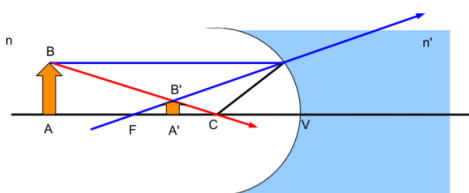
Si noti che l'immagine appare rovesciata. Vale inoltre

$$\frac{A'B'}{q - r} = \frac{AB}{p + r}$$

da cui definiamo m l'ingrandimento

$$m = \frac{A'B'}{AB}$$

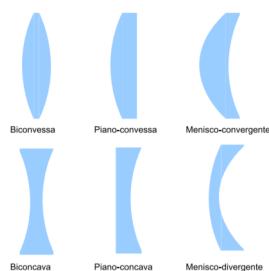
l'ingrandimento vale 1 quando $p = q - 2r$. Ripetiamo il procedimento per il diotro concavo. Tracciamo la retta parallela all'asse, quando incontra il diotro essa diverge, consideriamo quindi la retta di senso opposto verso F , tracciamo una seconda retta da B al centro del diotro. Il punto in cui le due rette si intersecano è l'immagine. L'immagine ottenuta è virtuale, dalla stessa parte dell'oggetto e dritta.



Lenti

Lenti sottili

dichiamo con il termine lente un sistema ottico costituito da materiale trasparente e omogeneo limitato da due superfici le quali separano un mezzo di indice di rifrazione n' da un mezzo di indice di rifrazione n . Una lente si dice sottile quando si può assumere che il suo spessore sia piccolo in confronto alle lunghezze in gioco, allora la lente si considera assimilata a un piano e si trascura ciò che avviene all'interno di essa quando è attraversata da raggi di luce. Esistono due famiglie principali di lenti: convergenti e divergenti.



Si dimostra che per le lenti sottili ciò che accade all'interno della lente è trascurabile e l'equazione dei punti coniugati per le lenti sottili è la seguente:

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f}$$

l'ingrandimento è dato da

$$m = \frac{q}{p}$$

Si definisce potere diottrico di una lente la quantità:

$$P = \frac{1}{f} = (n - 1) \left(\frac{1}{r_1} - \frac{1}{r_2} \right)$$

Specchi sferici

Una conseguenza di riflessione e rifrazione è che, tramite strumenti opportuni, è possibile deflettere raggi incidenti paralleli all'asse in modo da farli convergere nel fuoco, qui si concentrano grandi quantità di energia innescando talvolta la combustione (specchi ustori).