

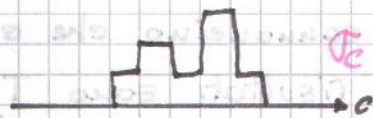
Esami: solo orale, no appelli \Rightarrow mettersi d'accordo

Le dispense contengono alcuni al **metodo di Montecarlo**, che però non verrà trattato a lezione.

- Problema: estrarre informazioni da grandi quantità di dati grezzi.

Supponiamo di voler misurare c . Avremo diversi risultati:

c_1, c_2, \dots, c_n



Vogliamo definire una funzione $\hat{c}(c_1 \dots c_n)$ che è la miglior stima che possiamo fare con questi dati.

La **precisione** ha a che fare con la larghezza della distribuzione. Possiamo definire l'intervallo di confidenza **CI** che la velocità della luce si trovi in un intervallo c_L, c_H come una probabilità:

$$P(c \in (c_L, c_H)) = CI \quad \leftarrow \text{confidence level}$$

$\{c_L, c_H\}$ si dice **intervallo di confidenza (IC)**.

Questo procedimento si dice **ricerca dei parametri**.

Le eq. di Maxwell ci dicono che $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$, che è un valore ben definito, fatto di grandezze misurabili con grande precisione.

\hat{c} è compatibile con questo valore?

Il fatto che una misura fornisca N valori diversi indica una aleatorietà intrinseca nel modo in cui è condotto l'esperimento.

Piccole variazioni al contorno possono produrre grandissimi scostamenti nel comportamento di un sistema. A volte le leggi deterministiche diventano caotiche (Poincaré) e devono essere trattate in modo diverso.

Consideriamo:

- eventi ripetibili
- condizioni dell'esperimento conosciute e costanti

Esempio: lanciamo una sfera monetina.

I possibili risultati sono T, C.

Definiamo successo l'evento T \Rightarrow num. di successi = m_T

$$f = \frac{m_T}{N}$$

(definizione di probabilità come frequenza)

Questo valore cambia se ripetiamo l'esperimento!

Ripetiamolo aumentando N :



All' aumentare di N , f si stabilizza su un valore.

Quindi come definizione possiamo usare:

$$P = \lim_{N \rightarrow \infty} \frac{m_T}{N}$$

Purtroppo, non c'è nessun modo per stabilire l'esistenza di questo

limite. Per ora assumiamolo in modo assiomatico.

Questo funziona solo con esperimenti RIPETITIVI.

Esistono altre definizioni.

Esempio: qual è la probabilità che il 200° numero dello sviluppo decimale di π sia 6?

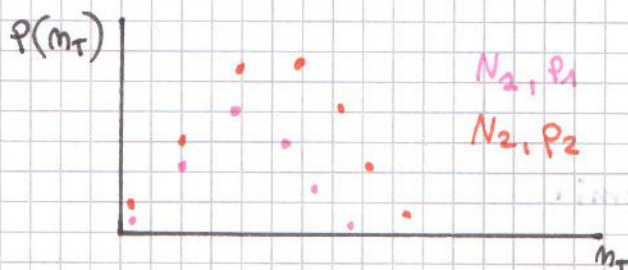
Non lo conosciamo: questo ci fa pensare ad $1/10$ (cifre da 0 a 9).

La teoria Bayesiana si fonda su questo ragionamento.

Probabilità e statistica sono due cose molto diverse.
 Consideriamo sempre il lancio di N monete.
 Se ho N lanci indipendenti:

$$P(m_T; N, p) = p^{m_T} (1-p)^{N-m_T} \binom{N}{m_T} \quad \left[\begin{array}{l} \text{problema} \\ \text{DEDUTTIVO} \end{array} \right]$$

con $p =$ probabilità di successo per il singolo lancio.
 Vedremo che questa formula non presenta ambiguità.



Supponiamo $N=100$, $m_T=22$. Quanto vale p ?
 Questo è un problema induttivo, non ha una soluzione sicura.

Possiamo però dare una stima \hat{p} di p , ad esempio in questo caso $\hat{p}=0.22$

La statistica è una scienza empirica basata sulla teoria delle probabilità.

Lancio di un dado: $\{1, 2, 3, 4, 5, 6\} \equiv S$

Cerchiamo di definire ogni evento elementare come elemento di un set.

Definiamo ad esempio l'evento "pari", che non è più elementare: $E_p \equiv \{2, 4, 6\}$ ($\bar{E}_p = E_p^c = \{1, 3, 5\}$)

In generale, $E \subseteq S$: un evento è sottoinsieme del set di possibilità.

- def: $A=B$ iff $x \in A \leftrightarrow x \in B$
- def: $A \supseteq B$ iff $x \in A \rightarrow x \in B$
- def: $A \cup B$ $x \in A \cup B$ iff $x \in A \vee x \in B$
- def: $A \cap B$ $x \in A \cap B$ iff $x \in A \wedge x \in B$

Queste def. stabiliscono un'ordine parziale negli insiemi che stiamo considerando.
Ci limiteremo a considerare insiemi limitati e quantità discrete.

Esercizio:

$$A \cup \emptyset = A$$

$$A \cap S = A \quad S \text{ insieme di tutti i sottoinsiemi.}$$

$$A \cup \bar{A} = S$$

$$\bar{\bar{A}} = A$$

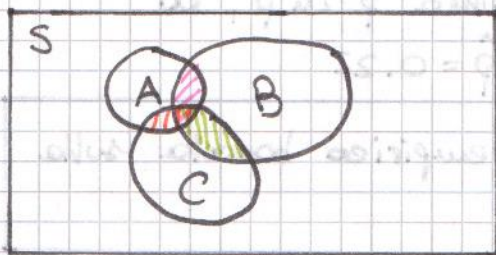
Consideriamo 3 insiemi.

$$A, B, C \subseteq S$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Possiamo convincerocene utilizzando una tavola delle verità.

Possiamo anche utilizzare: **diagrammi di Venn**.



Questo è il metodo più intuitivo.

Esercizio: verificare che $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

Regola di de Morgan

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

Un evento è un sottoinsieme del set elementare, i cui elementi sono mutualmente esclusivi.

Chiamiamo S lo **SPAZIO UNIVERSO**.

$E \subseteq S$; la classe degli eventi è la classe di tutti

i possibili sottoinsiemi di S .

Questo vale per quantità discrete.

La classe degli eventi è chiusa per le operazioni di unione ed intersezione.

$$C_E = \{S, \emptyset, A_1, A_2\}$$

Richiesta:

$$A_1 \in C_E \rightarrow \bar{A}_1 \in C_E$$

$$A_1 \in C_E \rightarrow A_1 \cup A_2 \in C_E$$

(...)

$$A_1, A_2, \dots, A_n \in C_E \rightarrow \bigcup_i A_i \in C_E$$

(con un numero numerabile di eventi)

Se questo è vero, allora C_E è una **sigma-algebra**.

Abbiamo una classe, ci serve una misura.

$P(E)$ probabilità matematica (misura)

- $P(S) = 1$

$$0 \leq P(E) \leq 1$$

- $P(E_1 \cup E_2) = P(E_1) + P(E_2) \underline{\underline{se}} E_1 \cap E_2 = \emptyset$

Queste tre proposizioni definiscono tutta la teoria della probabilità.

Come si sommano le probabilità?

Es.: 2 giocatori di tiro al piattello, con prob. di successo $p_1 = 80\%$ e $p_2 = 70\%$. Con che p. un piattello viene colpito da UNO dei due?

È una somma di eventi esclusivi $\rightarrow 94\%$

$$(= 80\% + 70\% \cdot (1 - 80\%))$$

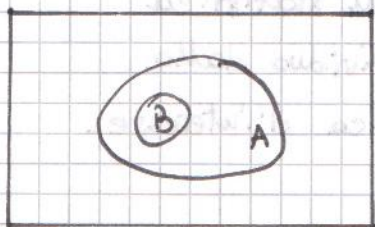
Esercizio

$$A \subseteq S \quad P(A)$$

$$P(\bar{A}) = 1 - P(A) \quad (\text{dim.})$$

$$P(\emptyset) = 0$$

$$B \subseteq A \rightarrow P(B) \leq P(A)$$



Dimostrazione:

$$A = A \cap (B \cup \bar{B}) = (A \cap B) \cup (A \cap \bar{B})$$

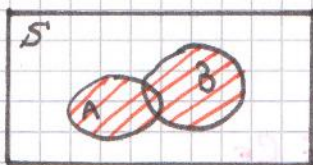
↑ esclusivi ↑

$$\Rightarrow P(A) = P(A \cap B) + P(A \cap \bar{B})$$

$$= P(B) + P(A \cap \bar{B}) \geq P(B)$$

Consideriamo: $A, B \subseteq S$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



(questo è il caso dei
tiratori al piattello)

Dim.: esercizio!

28/09/11

www.pi.infn.it/atlas → studenti → pdf lezioni / ps

Non sono aggiornatissime (anno scorso - alcuni errori).

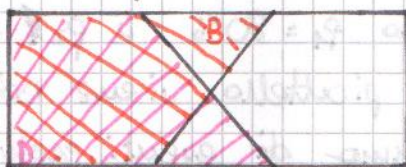
Abbiamo visto la definizione matematica di "probabilità" come
misura su uno spazio di campioni.

A volte si manifesta la necessità di imporre alcune condizioni
sull'esperimento, che si traducono in condizioni sulle
probabilità.

Es. di avere una città con N individui, N costante durante
la misura.

Vogliamo studiare il sottocampione di donne, ed il
sottocampione di persone con i capelli biondi.

$$\begin{cases} N_D \\ N_B \end{cases} \quad \begin{cases} P_D \equiv \frac{N_D}{N} \\ P_B \equiv \frac{N_B}{N} \end{cases}$$



P donne bionde? $P(D \cap B) = \frac{N_{D \cap B}}{N}$

definendo il campione come il numero di donne $P(B|D) = \frac{N_{D \cap B}}{N_D} = \frac{N_{D \cap B}}{N} \cdot \frac{N}{N_D} = \frac{P(D \cap B)}{P(D)}$

Si legge "probabilità di B dato D"

Il termine **popolazione** è usato molto spesso in statistica.
La "misura" consiste nell'estrarre un individuo dalla
popolazione, ed osservare la caratteristica d'interesse.

Le popolazioni possono essere binomiali, poissoniane, ecc.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{PROBABILITA' CONDIZIONALE}$$

Questo è il modo usato per aggiungere condizioni all'esperimento. Si verifica facilmente che:

- $0 < P(A|B) \leq 1$
- $P(S|B) = 1$
- $P(A \cup C|B) = \frac{P(A \cup C \cap B)}{P(B)} = \frac{P(A \cap B \cup C \cap B)}{P(B)}$
 $= \frac{P(A \cap B) + P(C \cap B)}{P(B)} = P(A|B) + P(C|B)$,
se $A \cap C = \emptyset$

Dimostrare che $P(\bar{A}|B) = 1 - P(A|B)$ (Esercizio).

Supponiamo che $P(A|B) = P(A)$, cioè che A e B non "comunichino". Questa è la definizione di **indipendenza**, dalla quale segue:

$$P(A \cap B) = P(A) \cdot P(B)$$

Supponiamo invece che $P(A|B) > P(A)$, o $P(A|B) < P(A)$. In questo caso gli eventi sono correlati. Esempio: malati di cuore / fumatori su una popolazione.

Esercizio: Sommosse in un carcere del centro america. Il direttore ordina di prendere 3 detenuti a caso ed ucciderne due. I tre sono James, William e Steve. James ha delle nozioni di statistica e chiede alla guardia (che dice sempre la verità) chi, tra W. ed S., sarà giustiziato. La guardia guarda il nome che appare primo sulla lista e dice "Steve". Che informazioni ne ricava J.? La lista è ordinata in ordine alfabetico.

Applicazione reale ad un giudizio forense. 1998 S. Clark fu accusata di aver ucciso il primo bambino a 2 sett. ed il secondo ad 8 settimane. La difesa: 2 casi di "SID" (sudden infant death). L'accusa interrogò un pediatra che dichiarò che questo può succedere con una probabilità di $1/100.000.000$. S.C. fu condannata su questa base.

$P(E|I) \approx 10^{-8} = P(E)$ (scorrelati)

↑ doppio SID imputata innocente

$P(I|E) ?$

I giudici hanno confuso queste due cose.

$P(E|I) = \frac{P(E \cap I)}{P(I)} \approx 10^{-8}; P(I|E) = \frac{P(E \cap I)}{P(E)}$

$\hookrightarrow = 1$ fino a prova contraria $\approx 10^{-8}$

I giudici avrebbero dovuto considerare, invece, $P(I|E)$.

Esercizio Una donna ha due figli. Uno dei due è un maschio. Qual è la probabilità che anche l'altro sia un maschio?

(MM), (MF), (FM) ← possibilità

\downarrow
 $1/3 = P$

A, B. $A \cap B = \emptyset$. I due eventi sono **dipendenti**, in quanto l'uno esclude l'altro.

$P(A) = 10\%$, $P(B) = 1$. Sono **dipendenti o indipendenti?**

Supponiamo di avere 3 eventi $A, B, C \subseteq S$.

$P(A \cap B \cap C) = P(A)P(B)P(C) ?$

Per l'indipendenza non vale la proprietà transitiva.

Esempio:

A \equiv dado 1 \rightarrow dispari

B \equiv dado 2 \rightarrow dispari

C \equiv dado 1 + dado 2 \rightarrow dispari

3 eventi sono correlati, perché se accadono A e B
 C non può verificarsi \Rightarrow dipendenza.
 Si verifica però che essi sono **a due a due**
indipendenti.

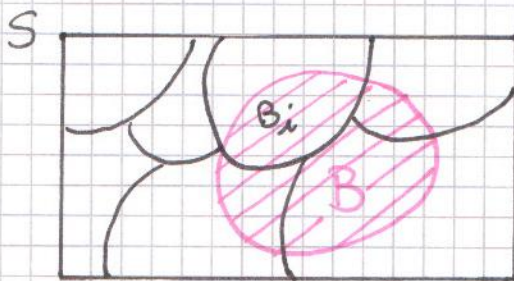
La formula di Bayes apre la porta alle probabilità soggettive.

$$A, B \subseteq S$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A|B) = P(B|A) \cdot \frac{P(A)}{P(B)}$$



$$\begin{aligned} \bigcup_i B_i &= S \\ B_i \cap B_j &= \emptyset \end{aligned}$$

Posso scrivere:

$$B = B \cap \left(\bigcup_i B_i \right) = \bigcup_i (B \cap B_i)$$

$$P(B) = \sum_i P(B|B_i) P(B_i) = \sum_i P(B \cap B_i)$$

Infatti, $P(A \cap B) = P(A|B) P(B) = P(B|A) P(A)$

Esercizio

3 scatole $(A_g, A_g), (A_g, A_u), (A_u, A_u)$

Estraendo una moneta d'oro da una delle scatole, dobbiamo trovare il modo di estrarre anche un'altra moneta d'oro. Usando il teorema di Bayes si riesce a trovare il modo.

29 settembre 2011

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Si suppone che $P(B)$ sia noto. Per simmetria: $P(B|A) = \frac{P(A \cap B)}{P(A)}$

$$\Rightarrow P(B|A) = \frac{P(A|B) P(B)}{P(A)}$$

PROBABILITÀ
A POSTERIORI

$P(B)$ - evidenza

$P(A|B)$ - verosimiglianza

$P(A)$ - probabilità a priori

La statistica Bayesiana è da un punto di vista pratico molto semplice da praticare.

Per poter usare la formula di Bayes conviene riscriverla.



S - spazio dei campioni

$$U_i A_i = S$$

$$A_i \cap A_j = \emptyset \text{ se } i \neq j$$

$$B = B \cap U_i A_i = U_i (B \cap A_i)$$

Abbiamo spezzato B in grandezze tra loro disgiunte

$$P(B) = \sum P(B \cap A_i) = \sum P(B|A_i) P(A_i)$$

$$P(A_i|B) = \frac{P(B|A_i) P(A_i)}{\sum P(B|A_j) P(A_j)}$$

Questa è una probabilità nel senso che è normalizzata a 1

Da un punto di vista della ^{probabilità} moderna ha poca importanza, ma è fondamentale per la probabilità di tipo bayesiano.

Cominciamo all'esempio di ieri delle monete.

Abbiamo estratto una moneta d'oro e vogliamo fare il modo di estrarre un'altra.

Siano U_1, U_2, U_3 le urne dove sono le monete

$P(U_i) = \frac{1}{3}$ La scelta iniziale dell'urna è a caso e quindi oggettiva ha la stessa probabilità.

Or si deve vedere come è fatta la verosimiglianza

$B \rightarrow G$ (gold)

$U_1 = A_u, A_u$

$U_2 = A_u, A_g$

$U_3 = A_g, A_g$

$A_1 \rightarrow M_1$

$$P(G|U_1) = 2$$

$$P(G|U_2) = \frac{1}{2}$$

$$P(G|U_3) = 0$$

$$P(U_1|G) = \frac{1 - \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} + 0} = \frac{2}{5}$$

La probabilità che l'urna da ho scelto abbia 2 monete d'oro è $\frac{2}{5}$ cioè è raddoppiata.

$P(U_2|G) = \frac{1}{3}$ Si ha a che fare con tutti eventi veri.

Se si dimostra che ^{se} questa procedura funziona allora sarebbe un metodo ideale per poter incrementare in modo controllato la conoscenza che si ha sul sistema.

Bayes funziona solo se conosciamo tutto il sistema

Esempio

Un'urna con 3 palle, non si conosce il colore ma solo o bianca o nera.

Se fo 3 estrazioni trovo sempre una palla bianca (tutte le volte ritiro la palla nell'urna). Posso usare Bayes per dire che ho 3 palle bianche?

H: l'urna ha 3 palle bianche

Io però non so se H è vero o no, siamo già fuori dalla statistica di Commodoro.

E: 3 palle bianche

E è l'evento vero, che è quello che io ho misurato

Bayes dice di costruire tante urne quante sono le combinazioni poi allora H diventa un evento.

U_1 (BBB)

U_2 (BBN)

U_3 (BNN)

U_4 (NNN)

$P(U_1) = 0$

Posso costruire una verosimiglianza cioè la probabilità dell'evento data l'urna.

Il problema è che se le palle non sono quantitative ma classiche allora non bastano 3 urne:

B B B

B N B

B B N

B N N

N N B

N B N

Sono tutti casi diversi:
"Palle" di Poisson

• Caso di urna quantitativa:

$$P(H|E) = \frac{P(E|H)P(H)}{\sum P(E|H)P(H)}$$

$$P(E|H) = 1$$

$$P(H) = \frac{1}{4}$$

Farlo a casa!!

Abbiamo visto due ci sono delle ambiguità nella scelta dei dati.

Esempio

La medicina usa la statistica nel caso della diagnostica computerizzata.

A_ positività del test

T_ risultato del test

Avremo solo 4 casi: $A\bar{A}$, $T\bar{T}$

$P(T|A) = 98\%$ passo il test e sono malato (sono malato | sono positivo)

$P(T|\bar{A}) = 3\%$ ~~passo il test ma non sono malato~~ (malato | test neg.)

quella che ci serve è $P(A|T)$ (positivo al test | malato)

$$P(A|T) = \frac{P(T|A)P(A)}{P(T|A)P(A) + P(T|\bar{A})P(\bar{A})}$$

Il problema è conoscere $P(A)$ che è la probabilità a priori, sap-

poniamo $P(A) = 10^{-3}$.

$$P(A|T) = \frac{0.98 \cdot 10^{-3}}{0.98 \cdot 10^{-3} + 0.03(1 - 10^{-3})} = 0.032$$

↳ sembra dire i tecnici dei metodi diagnostici sono sbagliati

Il problema è **prior**, il quale determina a priori la probabilità che sia malato; ma il sistema di diagnostica non funziona così, perché il medico ci deve mettere il suo **prior soggettivo**.

È inutile allora fare il test di screening generalizzato alla popolazione.

Il prior è imposto e non misurabile

Rifacendo di nuovo il test (lo stesso) allora uso il prior che ho stabilito prima, cioè il 3%, e allora se risulta ancora positivo la probabilità che sia davvero positivo è salita al 50%.

Esempio

Si costruisce un fascio e si trovano molte particelle π , e , μ , ma quelle che ci interessano sono solo gli elettroni.

Supponiamo che gli elettroni sono il 98%, e i π decimo lo stesso seguito con probabilità 3%.

(e) $P(\pi)$ è quello che si chiama prior e dipende dalla composizione del fascio.

Possiamo migliorare il trigger mettendo non 1 ma più sistemi

di trigger, in questo caso è impossibile evitare il teorema di Bayes.

Esempio

Ci sono 3 porte uguali solo una è quella giusta.

Viene poi scelto qual è una delle sbagliate, allora si dà la possibilità di cambiare la porta.

Questa cosa si può risolvere in due modi; o con Bayes o con un solo semplice ragionamento.

Fare a casa anche questo.

Esempio

Nel '48 un sottomarino non arrivò al porto, allora fu organizzata una ricerca. Un matematico ebbe l'idea di fare una ricerca bayesiana.

La ricerca fu fatta dividendo il mare in quadratini, ad ognuno fu dato una data probabilità: una che il sottomarino fosse naufragato e una legata alla probabilità che i mezzi lo ritrovassero.

La ricerca era iniziata e fatta considerando le zone via via più favorevoli.

$p = P(\text{relitto appartiene al quadratino})$

$q = P(\text{trovare relitto} | \text{relitto nel quadratino})$

$$p \rightarrow \frac{p(1-q)}{(1-q)p + (1-p)}$$

Il sottomarino si trovava nella regione prevista.

All'inizio del 1700 si era già data la probabilità di un evento binomiale.

Si fanno N lanci: $P(n | N, \theta) = \binom{N}{n} \theta^n (1-\theta)^{N-n}$

In generale siamo interessati ad avere n successi, per questo c'è il coefficiente binomiale davanti.

Vogliamo ora cercare $P(\theta | n, N) = \frac{P(n | N, \theta) P(\theta)}{\int_0^1 P(n | N, \theta) P(\theta) d\theta}$

Bayes ebbe dei dubbi su questa formula; noi sappiamo dare un significato a $P(n)$, ma non a $P(\theta)$; ma dato che non so niente di θ allora lo posso fare. (\rightarrow prob. unif. forme)

La cosa che non seppe fare è dare un significato a $P(\theta)$, cioè la probabilità a priori; la cosa più banale da fare è stato mettere $P(\theta) = \text{cost}$, ma questa scelta non è invariante per cambiamento di parametro.

$\theta_1, \theta_2, \dots, \theta_k$ non so niente di questi parametri.

$$P(\theta) = \frac{1}{k} \quad \forall$$

Si si accorge che l'unico parametro importante è θ_1 e degli altri non ci interessa niente

$$\phi = \phi_0 \quad \text{per } \theta = \theta_1$$

$$\phi_1 \quad \text{per } \theta \neq \theta_1$$

$$P(\phi_0) = \frac{1}{2} \quad \text{e} \quad p(\theta_1) = \frac{1}{k} \quad \text{e' inconsistente}$$

Per questo motivo Bayes non pubblicò il suo teorema.

Bayes disse: nel caso in cui non si sappia niente del parametro dobbiamo metterlo costante.

Esempio

Supponiamo di voler misurare il tempo di decadimento di una sostanza radioattiva.

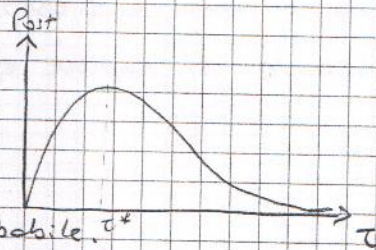
$$t = \frac{1}{2} e^{-t/\tau}$$

$$P(\tau) = \text{cost}$$

↳ Nel caso in cui non si sappia niente del radioattivo

Si fa la misura e si ottiene t_1

$$P(\tau | t_1) = \frac{P(t_1 | \tau) P(\tau)}{\int_0^{\infty} P(t_1 | \tau) P(\tau) d\tau} = \frac{\frac{1}{2} e^{-t_1/\tau} \cdot k}{\int_0^{\infty} \frac{1}{2} e^{-t_1/\tau} \cdot k d\tau}$$



Il massimo della funzione è il valore τ probabile, τ^*

Il massimo della probabilità a posteriori è un estimatore del parametro.

Ci possiamo chiedere cosa succede usando un'altra distribuzione per t , per esempio: $t = \lambda e^{-\lambda t}$

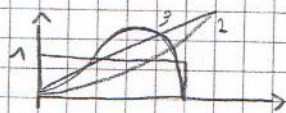
Se è stato fatto tutto giusto dovremmo trovare $\lambda^* = \frac{1}{\tau^*}$, fare da noi a caso il caso $P(\lambda) = \text{cost}$ e vedere se viene la stessa cosa, provare anche numericamente a stimare λ^* , proviamo anche a mettere

$P(\tau) = \frac{1}{\tau}$ e capire come mai.

$P(\tau) = \frac{1}{\tau} \iff P(\lambda) = \frac{1}{\lambda}$ questo prior da molta più consistenza.

Torniamo alla distribuzione binomiale, il prior è $P(\theta) = \text{cost}$, la zoba dell'integrale non ci interessa perché è solo un termine di non variazione.

$$P_0(\theta) = \text{cost} \quad N=0$$

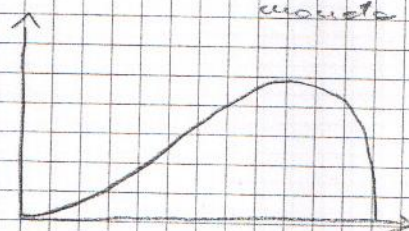


Esempio del lancio di una moneta

$$N=1 \quad P_1(\theta | T) \sim \theta$$

$$N=2 \quad P_2(\theta | TT) \sim \theta^2$$

$$N=3 \quad P_3(\theta | TTT) \sim \theta^3 (1-\theta)$$



Dopo N lanci si trova questo

Ad ogni passo prendo come prior quello che ho trovato al passo precedente

$$P_1(\theta|T) = \theta P_0(\theta)$$

$$P_2(\theta|T) = \theta \theta P_0(\theta)$$

All'aumentare di N l'assimilia della costante si perde.

Esercizio da fare

" T ": $\text{ran}(\cdot) > 0, \neq$ altrimenti "crace"

$f = \theta^T (1-\theta)^{1-T}$ questa è la funzione che vogliamo scrivere

$$I_T = 1 \text{ se } T \\ 0 \text{ se } C \Rightarrow f = \theta^{I_T} (1-\theta)^{1-I_T}$$

Provare a fare un disegno della funzione ad ogni passo, fatto un certo numero di passi $P_{\text{post}}(\theta) = \Pi \dots$ e trovare il massimo di questa funzione. Guardare come funziona nei due tipi di prior.

Il massimo è un indicatore del valore del parametro, quando diventa costante si capisce quando la teoria è consistente.

3/10/11

$$P(\mu|m) = \frac{P(m|\mu) P(\mu)}{\int_0^1 d\mu P(m|\mu) P(\mu)}$$

Per il prior si può utilizzare un valore convenzionale, però ha una certa aleatorietà.

In un esperimento generico abbiamo tanti risultati ("evento aleatorio"):

$$E: e_1, e_2, e_3, \dots, e_n$$

Vorremmo una funzione che trasferisca il risultato della misura in un indicatore numerico.

Ad es. per il lancio di una moneta:

$$\text{testa} \rightarrow 1$$

$$\text{croce} \rightarrow 0$$

$$E: e_1, \dots, e_n \rightarrow N: I_1, \dots, I_n$$

$$\bar{E}, \text{ in questo caso, } \sum_i I_i = \#T. \quad P(e=T) = p$$

Supponiamo che la nostra variabile aleatoria sia un numero reale distribuito tra $-\infty$ e $+\infty$.

Non si può parlare di probabilità che $e = x \in \mathbb{R}$, perché è praticamente nulla. Pensiamo quindi a:

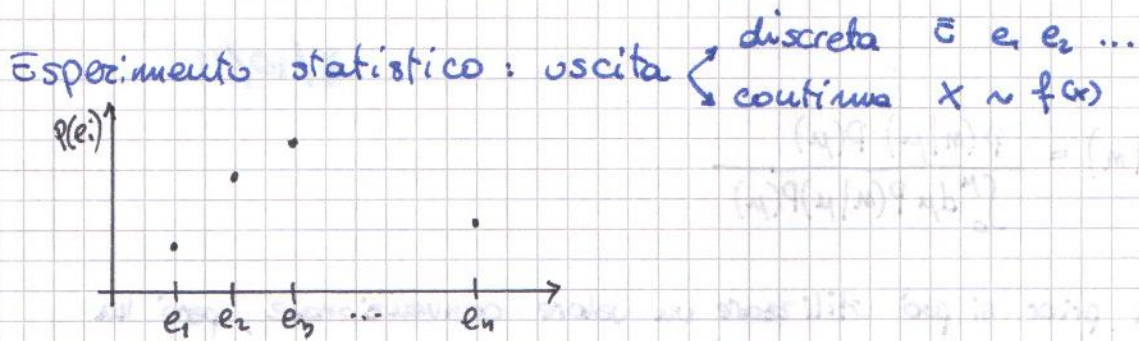
$$X \sim f(x) \quad (\text{la var. aleatoria } X \text{ è distribuita secondo una funzione } f)$$

$$P(X \in (a, b)) = \int_a^b f(x) dx$$

$f(x)$ è una DENSITA' DI PROBABILITA'.

$$P(X \in (x, x+dx)) = f(x) dx$$

Lanciamo N monete - la probabilità di ottenere n teste è binomiale.



Ci interessano il baricentro e la lunghezza di questa distribuzione, piuttosto che i singoli valori.

Valore d'aspettazione $E(x) = \sum_i x_i P(X=x_i)$

Questo corrisponde alla media della distribuzione.

Si verifica che:

$$E(x+y) = E(x) + E(y)$$

$$E(kx) = k E(x)$$

Momento n -esimo della distribuzione

$$\mu_n = E(x^n) = \sum_i x_i^n P(x_i)$$

Il secondo momento che ci interessa è la varianza.

Varianza $Var(x) = \sum_i P(x_i) (x_i - E(x))^2 = E\left(\left(x - E(x)\right)^2\right) = \bar{x}^2 - \bar{x}^2$

Essa è una misura della dispersione dei punti attorno al valore d'aspettazione:

$$\sigma = \pm \sqrt{Var(x)}$$

Supponiamo di avere:

$$z = x + y$$

Se x e y sono indipendenti:

$$P(x|y) = P(x)$$

$$P(x \cap y) = P(x) P(y)$$

Ovvero:

$$"x" \equiv X = x_i$$

$$P(X=x_i \cap Y=y_j) = f(x_i, y_j)$$

$$E(z) = \sum_i (x_i + y_i) P(x_i, y_i) = \sum_i x_i P(x_i, y_i) + \sum_i y_i P(x_i, y_i) \\ = \sum_i x_i \sum_j P(x_i, y_j) + \dots$$

Si dimostra che $\sum_j P(x_i, y_j) = P(x = x_i)$ - Infatti:

$$x = x_i \cap U_i (y = y_i) = U_j (x = x_i \cap y = y_j)$$

$$P(x = x_i) = P(U_i (x = x_i) \cap (y = y_i)) = \\ = \sum_j P(x = x_i \cap y = y_j) = \sum_j P(x = x_i | y = y_j) P(y_j) \\ = \sum_j P(x_j, y_j)$$

In definitiva, se $z = \sum_i x_i$ allora $E(z) = \sum_i E(x_i)$.

Possiamo al secondo valore di aspettazione.

$$\text{Var}(x) = E((x - \mu)^2) = E(x^2 - 2\mu x + \mu^2) = \\ = E(x^2) - 2\mu^2 + \mu^2 = E(x^2) - \mu^2$$

Se abbiamo due variabili aleatorie:

$$\text{Var}(x+y) = E((x+y - (\mu_x + \mu_y))^2) = \\ = E((x - \mu_x)^2) + E((y - \mu_y)^2) + 2E((x - \mu_x)(y - \mu_y)) \\ = \text{Var}(x) + \text{Var}(y) + 2\text{Cov}(x, y)$$

Covarianza di x ed y

$$\text{Cov}(x, y) = E((x - \mu_y)(y - \mu_x))$$

Possiamo esplicitare il prodotto e scrivere:

$$\text{Cov}(x, y) = E(x \cdot y) - \mu_x \mu_y$$

Si verifica che se gli eventi x ed y sono indipendenti la loro covarianza è nulla.

$$P(x|y) = P(x) \quad P(x \cap y) = P(x)P(y)$$

Nel linguaggio delle distribuzioni:

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

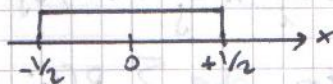
$$P(x = x_i, y = y_i) = P(x = x_i)P(y = y_i)$$

proseguire la dimostrazione.

Esempio. L'evento A ha una distribuzione $X \sim U(-\frac{1}{2}, +\frac{1}{2})$
 L'evento B " " " " X^2

$$E(A) = 0$$

$$\begin{aligned} \text{Cov}(A, B) &= E(B \cdot A) - 0 \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} x^3 dx - 0 = 0 \end{aligned}$$



Esercizio: dimostrare che

- $E(x+y) = E(x) + E(y)$
- $E(c_1 x + c_2 y) = c_1 E(x) + c_2 E(y)$
- $V(x+y) = V(x) + V(y) + 2 \text{Cov}(x, y)$
- $V(c_1 x + c_2 y) = c_1^2 V(x) + c_2^2 V(y) + 2c_1 c_2 \text{Cov}(x, y)$
- $\text{Cov}(x+y, z) = \text{Cov}(x, z) + \text{Cov}(y, z)$
- x_i scorrelate $\Rightarrow V(\sum_{i=1}^n x_i) = nV(x_i)$

Funzione generatrice dei momenti:

$$\phi_x(t) = E(e^{tx}) \Big|_{t=0}$$

È molto simile alle trasformate di Fourier, ma rimane nel campo reale.

$$\begin{aligned} &\approx E\left(1 + tx + \frac{1}{2}(tx)^2 + \dots\right) = \\ &= 1 + tE(x) + \frac{t^2}{2}E(x^2) + \dots = \sum_{k=0}^{\infty} \frac{t^k E(x^k)}{k!} = \sum_{k=0}^{\infty} \frac{t^k \mu_k}{k!} \end{aligned}$$

Per ottenere μ_k posso calcolare:

$$\mu_k = \left. \frac{\partial^k \phi_x(t)}{\partial t^k} \right|_{t=0}$$

Si ha inoltre: $z = x+y$, y, x indep.

$$\phi_z(t) = E(e^{t(x+y)}) = E(e^{tx}) E(e^{ty}) = \phi_x(t) \phi_y(t)$$

Teorema $f_1, f_2 : \mu_k^1 = \mu_k^2 \forall k$
 $f_1 - f_2 = \sum C_m x^m \Rightarrow f_1$ ed f_2 sono identiche

Dimostrazione:

$$0 \leq \int_{-\infty}^{+\infty} (f_1 - f_2)^2 dx = \int_{-\infty}^{+\infty} (f_1 - f_2) \sum C_m x^m dx =$$

$$= \sum C_m \int_{-\infty}^{+\infty} (x^m f_1 - x^m f_2) dx = 0$$

Le trasformate di Fourier hanno il problema che non sempre l'inversa è definita.

$$N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \phi_x(t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x^2}{2} - tx\right)} dx =$$

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

$$x, y \sim N(0,1) \Rightarrow \phi_{x+y} = e^{t^2}$$

Supponiamo di avere:

$$\phi_{x \cdot a}(t) = E(e^{t \cdot x \cdot a}) = \phi_x(at)$$

$$x \sim N(0,1)$$

$$\sigma \cdot x \sim N(0,2)$$

$$z \equiv \sqrt{2} \cdot x \quad z \sim N(0,2) \quad \phi_z(t) = e^{t^2}$$

$\Rightarrow x+y$ è distribuito come una variabile normale (Gaussiana) con varianza = 2.

Esempio di Bernoulli: lancio di una moneta, che ha prob. di successo p .

$$\phi_{\text{Bern.}}(t) = E(e^{tx}) = pe^t + (1-p)$$

$$E(x) = \frac{\partial \phi}{\partial t} \Big|_{t=0} = p$$

$$E(x^2) = \frac{\partial^2 \phi}{\partial t^2} \Big|_{t=0} = p$$

$$\Rightarrow \text{Var}(x) = E(x^2) - E(x)^2 = p(1-p)$$

$$y = \sum x_i \quad \phi_y(t) = (pe^t + (1-p))^N \quad E(y) = Np$$

Esercizio calcolare $Var(y) = Np(1-p)$

Esercizio

$$x \sim U(-\frac{1}{2}, \frac{1}{2}) \quad \left(x = \begin{cases} 1 & x \in (-\frac{1}{2}, \frac{1}{2}) \\ 0 & \text{fuori} \end{cases} \right)$$

$$\phi_x(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{xt} dx = ?$$

$$Var(x) = \frac{1}{12}$$

5/10/11

Dispense aggiornate sul sito -

Esercizi sulla covarianza

• $x, y \sim N(\mu, 1)$

$$U = X + a \quad V = Y + x$$

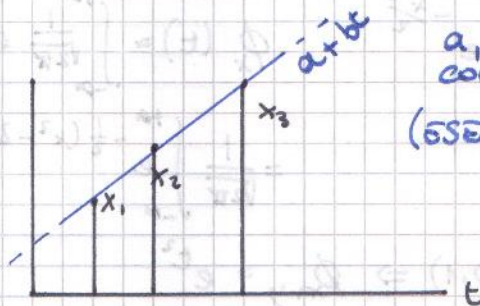
• x, y iid $\sim U(0, 1)$

$$\begin{cases} u = x \\ v = u^2 + y \end{cases} \quad Cov(u, v)$$

• $\varphi \sim U(0, 2\pi)$

$$\begin{cases} x = \cos \varphi \\ y = \sin \varphi \end{cases}$$

correlate?



$$\rho(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x) Var(y)}}$$

Si dimostra che ρ è una variabile normalizzata compresa fra -1 e 1.

Abbiamo parlato di processi di branching -

Supponiamo di avere

x_i iid

$$Y = \sum_{i=1}^N x_i$$

N aleatoria

identicamente distribuite ed indipendenti

$$E(Y) = \sum_j j P(Y=j)$$

unione di eventi disgiunti

Possiamo scrivere: $(Y=j) = (Y=j) \cap (U_m N = m)$

Questo si dice **CONDITIONING**.

$$P(Y=j) = \sum_m P(Y=j \cap N=m) = \sum_m P(Y=j | N=m) P(N=m)$$

$$= \sum_m P(m) P\left(\sum_{i=1}^m x_i = j\right) \quad \text{dove } m \text{ è un numero}$$

e non una variabile aleatoria.

$$= \sum_m P(m) E(Y_m) \quad \text{con } E(Y_m) = P\left(\sum_{i=1}^m x_i = j\right)$$

$$= \sum_m P(m) m E(x) = E(m) E(x)$$

È anche abbastanza intuitivo.

Esercizio

Calcolare con questo stesso metodo la varianza di y (sapendo che $V(y) = E(y^2) - E(y)^2$)

Il precedente è un tipico esempio di processo di **branching**: un processo che genera le x_i ed uno che governa i valori di N .

Esempio: $N \sim$ numero di clienti alla cassa / st
 $x_i \sim$ tempo trascorso alla cassa t cliente (in media)

Esempio: rivelatore di luce

fascio di fotoni \rightarrow rivelatore

$P(m_y)$
 numero di fotoni

Prima del rivelatore mette un filtro che ne lascia passare il 10%...

Per risolvere questi problemi sono utili le **funzioni generatrici di probabilità**.

$$x \equiv 0, 1, 2, \dots$$

$$\pi(s) = E(s^{x_i}) = \sum s^{x_i} P(x_i)$$

Proprietà

- $\pi(1) = 1$
- $\left. \frac{\partial \pi}{\partial s} \right|_{s=1} = \sum x_i P(x_i) = E(x)$
- $\left. \frac{\partial^2 \pi}{\partial s^2} \right|_{s=1} = E(x^2) - E(x)^2$

Se ho x, y indipendenti: $\pi_{x+y}(s) = E(s^{x+y}) = \pi_x(s) \pi_y(s)$

Esempio $A = \{a_0, a_1, a_2\}$ $B = \{b_0, b_1, b_2\}$ $a_0 = P(A=0) \dots$

$$\pi_A(s) = a_0 + a_1 s + a_2 s^2$$

$$\pi_{A+B}(s) = (a_0 + a_1 s + a_2 s^2)(b_0 + b_1 s + b_2 s^2) = \sum_r c_r s^r$$

$$\text{con } cr = \sum_{i+j=r} a_i b_j = \sum_i a_i b_{r-i} \quad (\text{sto facendo una convoluzione})$$

Caso realistico: lancio un dado e voglio vedere il risultato.

$$\mathcal{TL}_{1D}(s) = \frac{1}{6} (s + s^2 + s^3 + s^4 + s^5 + s^6)$$

Supponiamo di voler lanciare 10 dadi e calcolare la probabilità che il risultato sia 50.

Occorre calcolare $\mathcal{TL}_{10D} = (\mathcal{TL}_{1D}(s))^{10}$ e prendere il coefficiente del termine $s^{50} \Rightarrow$ computer.

$$S \equiv \sum_{i=1}^N x_i \quad N \text{ aleatoria}$$

• calcolare $P(S) = \sum_n P(S=j | N=n) P(n)$

• calcolare $\mathcal{TL}_S(t) = \sum_k t^k P(S=k) =$

$$= \sum_n P(n) \underbrace{\sum_k t^k P\left(\sum_{i=1}^n x_i = k\right)}_{(\mathcal{TL}_x(t))^n} = \sum_n P(n) (\mathcal{TL}_x(t))^n$$

$$= \mathcal{TL}_n(\mathcal{TL}_x(t))$$

Questi metodi risalgono alla scuola di Eulero.

6/10/11

$$S = \sum_{i=1}^N x_i \quad x_i, N \text{ variabili aleatorie}$$

Se il problema si presta ad essere trattato in questo modo, (cioè i valori sono discreti e tutti > 0) si definisce:

$$T_X(s) = E(s^x) = \sum_x s^x p(x)$$

Ad es. per la binomiale:

$$T_B(s) = sp + (1-p)s^0$$

Mentre se abbiamo una distribuzione poissoniana: $p(r) = \frac{e^{-\mu} \mu^r}{r!}$

$r = 0 \dots \infty$

$$T_P(s) = \sum_{r=0}^{\infty} s^r \frac{e^{-\mu} \mu^r}{r!} = e^{-\mu} e^{\mu s} = e^{\mu(s-1)}$$

E quindi:

$$E(x^n) = \left. \frac{\partial^n T_X(s)}{\partial s^n} \right|_{s=1}$$

mentre

$$p(r) = \left. \frac{\partial^r T(s)}{\partial s^r} \right|_{s=0}$$

∃ software per calcolare analiticamente questo tipo di probabilità.

Esercizio

Un numero di fotoni, $n \sim \mathcal{D}(\mu)$, viene fatto passare attraverso un filtro, che ne fa passare un certo numero: ϵn

$$\begin{aligned} T_{\text{fotoni}}(s) &= T_{\text{beam}}(T_{\text{filtro}}(s)) = (1-p) + p[(1-\epsilon) + \epsilon s] \\ &= (1-p\epsilon) + \epsilon p s \end{aligned}$$

⇒ la distribuzione è ancora binomiale, con probabilità ϵp .

E se inizialmente il fascio era poissoniano?

$$= e^{\mu(\epsilon s + (1-\epsilon) - 1)} = e^{\epsilon \mu (s-1)}$$

è ancora poissoniana, con media $\epsilon \mu$.

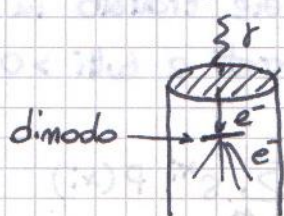
Esercizio

Allo cassa del supermercato:

$$\pi_{\text{cassa}}(s) = e^{\mu(e^{\lambda(s-1)} - 1)}$$

Poissoniana composta

La Poissoniana composta funziona anche con problemi di moltiplicazione. Es.: fotomoltiplicatori:



m fotoni incidenti \sim Poissoniana

m_e elettroni prodotti \sim Poissoniana

m elettroni secondari \forall elettrone primario

Nei fotomoltiplicatori reali i dinodi sono molti.

DISTRIBUZIONE DI BERNOULLI

Processo stocastico con variabile aleatoria che può ottenere solo 2 valori. Ogni esperimento non "ricorda" i lanci precedenti.

- stazionarietà
- costanza dei parametri

$$X = \begin{matrix} T \\ C \end{matrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$P(T) = p \quad P(C) = 1 - p$$

$$E(X=1) = p \quad \text{Var}(X) = p(1-p)$$

DISTRIBUZIONE BINOMIALE

$\equiv N$ processi di Bernoulli indipendenti.

Esempio: si osserva una catena temporale di 20 "C".

Si deve scommettere. Si punta su T o C?

$$P(T | m \text{ "C" precedenti}) = \frac{P(T \cap m \text{ C})}{P(m \text{ C})} = P(T)$$

La distribuzione non ha memoria degli eventi precedenti.

Lancio una moneta n volte \Rightarrow lo spazio dei campioni contiene 2^n risultati. Ogni evento (compreso n volte T) ha esattamente la stessa probabilità.

L'indipendenza stocastica è una cosa ben verificata.

Esercizio 10 operai usano ogni tanto, per 12 minuti all'ora, una certa potenza elettrica, la quale però è progettata t.c. al max 6 operai la possono utilizzare contemporaneamente. Calcolare $P(\text{overload})$.

STATISTICA GEOMETRICA

n = numero delle prove fino al primo successo -

$$P(n) = p(1-p)^{n-1}$$

Gioco di St. Petersburg: \times guadagnare sempre in un gioco d'azzardo. Auto 1€ , gioco, raddoppio, mi fermo al primo successo. Ho puntato c^{n-1} e vinto $c^n \Rightarrow$ il guadagno è c .

$$\begin{aligned} \text{gioco: } c + 2c + 4c + \dots + 2^{n-1}c &= c \cdot \frac{2^n - 1}{2 - 1} = c(2^n - 1) \\ \text{riuso: } 2^n c & \end{aligned}$$

Il problema è che bisogna essere abbastanza ricchi da sostenere le perdite fino al guadagno.

$$E(L) = E(c(2^n - 1)) = ?$$

Bisogna che il gioco abbia $p > \frac{1}{2}$. Già a $p = \frac{1}{2}$ la serie diverge e non ho speranza di vittoria.

Esercizio probabilità p che una radio si rompa -

$P(k)$ che funzioni per k giorni? \bar{X} una distrib. geometrica -

$$P(\text{rottura al giorno } k) = p(1-p)^{k-1}$$

$$E(k) = ? \quad \text{Var}(k) = ?$$

$$\begin{aligned} E(k) &= \sum_{k=0}^{\infty} k p(1-p)^{k-1} = p \sum_{k=0}^{\infty} k (1-p)^{k-1} = -p \sum_{k=0}^{\infty} \frac{\partial}{\partial p} (1-p)^k \\ &= -p \frac{\partial}{\partial p} \sum_{k=0}^{\infty} (1-p)^k = -p \frac{\partial}{\partial p} \left(\frac{1}{1-(1-p)} \right) = p \frac{1}{p^2} = \frac{1}{p} \end{aligned}$$

Supponiamo $p = 10^{-2} \Rightarrow$ in medio la radio funziona per 100 giorni.

Qual è la probabilità che la radio funzioni per, ad esempio, 75 giorni?

$$P(n < M) = ?$$

Vedremo che la media non dà informazioni affatto precise sulla probabilità di funzionamento della radio. Allora se ne compra una seconda di scorta.

$$Z = X + Y \quad \text{tempo } \times \text{ cui una delle due radio funziona -}$$

$$P(z=c) = \sum_j P(x=a_i \cap Y=b_j), \quad a_i + b_j = c$$

$$= \sum_i P(x=a_i \cap Y=c-a_i) =$$

$$= \sum_j P(x=a_j) P(Y=c-a_j)$$

$$a_j \rightarrow l \in \mathbb{R}, < \infty$$

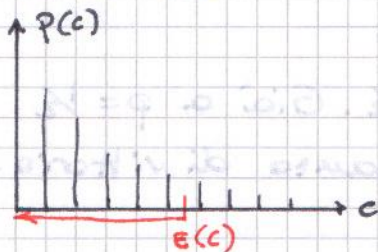
$$= \sum_{l=1}^{c-1} P(x=l) P(Y=c-l)$$

È una relazione di convoluzione $\left(\int f(x) g(z-x) dx \right)$

$$= \sum_1^{c-1} p(1-p)^{c-l-1} p(1-p)^{l-1} = p^2 (1-p)^{c-2} (c-1)$$

Questa non è più una probabilità geometrica.

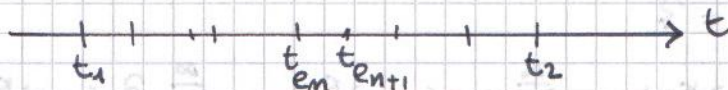
$$p_1 = \frac{1}{10}, \quad p_2 = \frac{1}{100}, \quad E(K) = ? \quad P(K > c) = ? \quad P(K < c) = ?$$



Da X, Y indipendenti abbiamo considerato la distribuzione di $Z = X + Y$, ed abbiamo visto che corrisponde alla distribuzione della convoluzione di X con Y .

DISTRIBUZIONE DI POISSON

Stessi postulati della prova di Bernoulli.



Es.: eventi che hanno probabilità sui fornes nel tempo di accadere

(come macchine che passano a un incrocio)

$$t_{m+1} - t_m \equiv \Delta t \sim e^{-t/\tau} \quad \text{numero di suddivisioni di } T$$

$$T \equiv t_2 - t_1 \quad P(m | pN) = \binom{N}{m} p^m (1-p)^{N-m}$$

Se $N \rightarrow \infty, p \rightarrow 0$, la distribuzione è binomiale

(difficile che ci siano due o più eventi in un $dt = \frac{t_2 - t_1}{N}$)

$$Np \equiv \mu \Rightarrow P(m | pN) = \frac{N!}{m!(N-m)!} \frac{\mu^m}{N^m} \left(1 - \frac{\mu}{N}\right)^N / \left(1 - \frac{\mu}{N}\right)^m$$

$$\xrightarrow{N \rightarrow \infty} \frac{e^{-\mu} \mu^m}{m!} \left(\frac{N(N-1) \dots (N-m-1)}{N^m} \right) \rightarrow \frac{e^{-\mu} \mu^m}{m!}$$

Ad es.:

$$B\left(k, \frac{1}{300} = p, N=300\right) = ? \quad B(k=0) = (1-p)^{300} = 0.254$$

Ma fare questo calcolo è difficile anche per un computer. Se invece usiamo la formula di Poisson con $r=0$, $\lambda=Np$ otteniamo $P(k=0)=.254$ con molta meno fatica.

11/10/11

Poisson \rightarrow probabilità di avere un certo conteggio in un esperimento dove il numero medio di conteggi è costante.

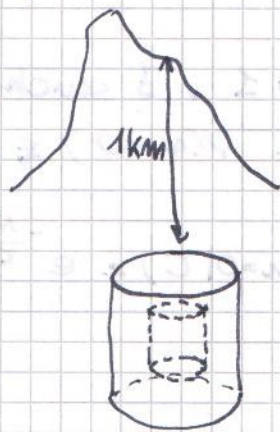
Esperimento: misura della vita media del protone.

Limite inferiore: età dell'universo (il numero barionico si conserva).

Teorie oltre il modello standard tipo SU(5) avevano predetto

$$\tau_{\text{protone}} \sim 10^{33} \text{ anni} \quad p \rightarrow e^+ \pi_0 \rightarrow \gamma\gamma$$

Gli e^- producono radiazione Cerenkov nel rivelatore.



Coverno con $V = 5 \cdot 10^7 \text{ m}^3$, schermato dai raggi cosmici, continuamente riempita di acqua purissima.

La superficie del cilindro è tappezzata di oltre 10k fotomoltiplicatori, sensibili al singolo fotone.

La regione interna "di Pdvcaia" ha $V = 5 \cdot 10^4 \text{ m}^3$.

$$N_p = \underbrace{\rho \frac{V}{A}}_{\# \text{ moli}} N_A \cdot (2 \div 18)$$

(H_2O : O contiene 8 p^+ , ma sono legati, non si sa se possono decadere)

$$A = 16 \quad Z = 2 \quad N_A = 6 \cdot 10^{23} \text{ mol}^{-1} \quad \Rightarrow N_p \approx 10^{34} \equiv N_0$$

Assumiamo che tutti questi protoni decidono indipendentemente se decadere o no.

$$N(t) = N_0 e^{-t/\tau}$$

$$\bar{N} = \frac{dN}{dt} \Delta t \approx N_0 \frac{1}{\tau} \Delta t \quad \text{numero medio di eventi attesi:}$$

\Rightarrow ci aspettiamo 10 eventi in un anno.

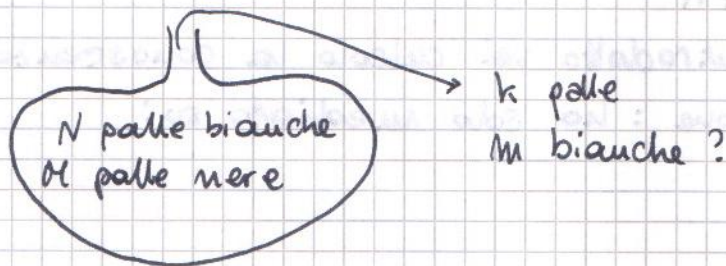
Questi 10 eventi devono essere distribuiti come uno

Distribuzione ipergeometrica

Il modello è di estrazione senza rimpiazzo.

Supponiamo di avere un'urna contenente N palle bianche ed M palle nere. Si estrae una palla alla volta e non si rimette dentro.

Anzi, estraiamo k palle e calcoliamo la probabilità che tra queste k ce ne siano m bianche.



Numero di modi con cui posso prendere m palle tra le N :

$$N(m_B | N_B) = \binom{N}{m}$$

$$N(k-m | M) = \binom{M}{k-m}$$

In quanti modi posso prendere k palle da tutte quelle che ho nell'urna?

$$N(k | N+M) = \binom{N+M}{k}$$

$$P(k | N, M) = \frac{\binom{N}{m} \binom{M}{k-m}}{\binom{N+M}{k}}$$

Probabilità ipergeometrica

Queste non sono sicuramente prove di Bernoulli: ogni estrazione modifica l'esperimento.

$E_i \equiv$ ho estratto (i) B $\rightarrow w_i \equiv 1$ se B
 0 se N

$$P(w_1 = 1) = \frac{N}{N+M}$$

In $P(w_2 = 1)$ possono cambiare sia numeratore che denominatore!

$$\begin{aligned}
 P(w_2=1) &= P(w_2=1 \cap \{w_1=0 \cup w_1=1\}) = \\
 &= P((w_2=1 \cap w_1=0) \cup (w_2=1 \cap w_1=1)) = \\
 &= P(w_2=1 \cap w_1=0) + P(w_2=1 \cap w_1=1) = \\
 &= P(w_2=1 | w_1=1)P(w_1=1) + P(w_2=1 | w_1=0)P(w_1=0) = \\
 &= \frac{N-1}{N+M-1} \cdot \frac{N}{N+M} + \frac{N}{N+M-1} \cdot \frac{M}{N+M}
 \end{aligned}$$

Facendo tutti i conti otteniamo:

$$P(w_2=1) = \frac{N}{N+M} \quad \text{come per la 1ª estrazione.}$$

In fatti non ho mai introdotto nel calcolo la conoscenza della precedente estrazione: ho solo mediato su i possibili risultati.

$$x_k \equiv \sum_{i=1}^k w_i$$

$E(x_k)$ è il valor medio di bianche che ho su k palline estratte.

$$E(x_k) = \sum_{i=1}^k E(w_i) = k \frac{N}{N+M}$$

L'unica cosa complicata è il calcolo di $\text{Var}(x_k)$: le x_i sono tutte correlate!

$$\text{Var}(x_k) = \sum_{i=1}^k \text{Var}(w_i) + 2 \sum_{i=1}^k \sum_{j>i}^k \text{Cov}(w_i, w_j)$$

$$\text{Var}(w_i) = P(w_i) (1 - P(w_i)) = \frac{N}{N+M} \left(1 - \frac{N}{N+M}\right)$$

$$\text{Cov}(w_i, w_j) = E(w_i \cdot w_j) - E(w_i)E(w_j)$$

Il risultato è:

$$\text{Var}(x_k) = k \frac{NM}{(N+M)^2} \left(1 - \frac{k-1}{N+M-1}\right)$$

Correzione per popolazioni finite.

Per $N, M \rightarrow \infty$ la varianza è quella binomiale.

Esempio: contare i cervi in un bosco.



Una mattina faccio un'incursione nel bosco, catturo e marcho n cervi.

Adesso la popolazione è data da n più $N - n$ animali.

Attendo che i cervi marcati si randomizzino. Poi ne catturo m (tra gli N totali) e ne trovo i marcati.

$$E(i) = \frac{rM}{N}$$

Se assumo che " i " sia rappresentativo di tutte le possibili catture ($E(x) \equiv i$):

$$N = \frac{m \cdot r}{i}$$

C'è lo stesso problema dei protoni: e se $i=0$?
Ho scelto m, r troppo piccoli rispetto ad N .

Possiamo alle distribuzioni continue.

Distribuzione uniforme $f(x) = \frac{1}{b-a}$ $a \leq x \leq b$

A simple graph showing a horizontal line segment from $x=a$ to $x=b$ on the x-axis. The y-axis is not explicitly labeled, but the function value is constant at $\frac{1}{b-a}$.

Distribuzione dei tempi: (= degli intervalli che intercorrono tra due eventi casuali consecutivi), con $P(t, t+dt) = \mu dt$

$$\bullet f(t) dt = P_{0|s}(0 | t, \mu) P(1 \text{ ev in } dt) = e^{-\mu t} \mu dt$$

$$\bullet P(T=t+dt) = P(T=t) (1 - \mu dt) \Rightarrow \frac{P(t+dt) - P(t)}{dt} = -\mu P(t)$$

nessun evento in $(t, t+dt)$

(Stiamo considerando eventi distribuiti poissonianamente).

Distribuzione esponenziale $P(t) = \mu e^{-\mu t}$

C'è una proprietà interessante:

$$P(t > a+b | t > a)$$

è indipendente da a .

RENEWAL

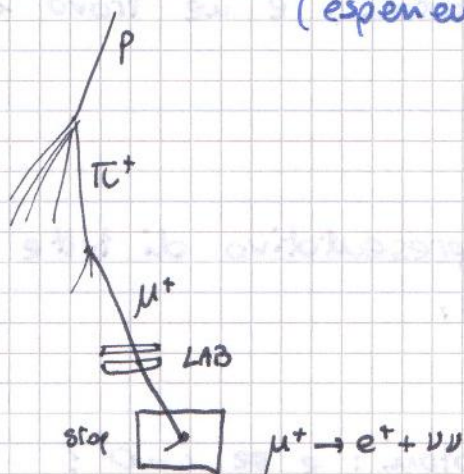
In fatti:

$$P(t > a+b | t > a) = \frac{P(t > a+b \cap t > a)}{P(t > a)} = \frac{P(t > a+b)}{P(t > a)}$$

$$P(t > x) = \int_x^{\infty} \mu e^{-\mu t} dt = e^{-\mu x}$$

$$\Rightarrow P(t > a+b | t > a) = \frac{e^{-\mu(a+b)}}{e^{-\mu a}} = e^{-\mu b}$$

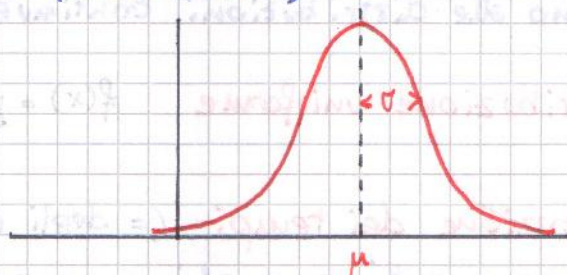
Applicazioni ad esempio nella misura dello vite media del μ - (esperienza laboratorio IF)



Distribuzione Gaussiana (NORMALE)

È una distribuzione limite (asintotico)

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



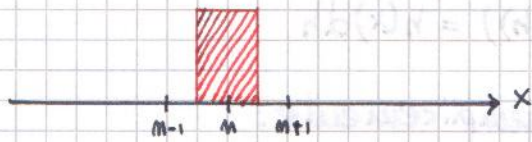
Se considero un set di x_i distribuite in qualsiasi modo, $S \equiv \frac{1}{n} \sum_{i=1}^n x_i$ è distribuita in modo Gaussiano.

Questa distribuzione fu ricavata indipendentemente da Gauss e Demoisire, il quale considerò una distribuzione binomiale di eventi $P(k, p, N)$ con $N \rightarrow \infty$, $p \rightarrow 0$ e $Np = \text{cost.}$

Consideriamo $N = 2M+1$, $p = 1/2 \Rightarrow$ la binomiale tende ad una Gaussiana.

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

Si è usato: $E(k) = pN \equiv \mu$, $\text{Var}(k) = 2N(1-p)p = \frac{N}{2} \equiv \sigma$



Mel discreto l'evento è " $x=m$ ";
 mel continuo " $x \in (m-1/2, m+1/2)$ "

$$P(\epsilon) = \int_{m-1/2}^{m+1/2} N(\mu, \sigma) dx$$

Se invece vogliamo $P(x > m)$: $P(x > m) = \int_{m+1/2}^{\infty} N(\mu, \sigma) dx$

Correzione necessaria per passare da discreto
 a continuo

Le probabilità non si sommano / trasformano come
 variabili reali, perché si tratta di misure. Non hanno,
 ad esempio, un verso o una direzione.

12/10/11

Abbiamo concluso le distribuzioni. Ora vediamo come
 utilizzarle.

CAMBAMENTO DI VARIABILE

$$X = x_1 \dots x_k$$

Supponiamo di essere interessati ad

$$h(x) = h_1, \dots, h_k$$

Le probabilità degli x_i sono conosciute. Per definizione:

$$P(h(x) = h_0) = \sum_{h(x_i) = h_0} P(x = x_i)$$

Esempio: $X = -\frac{\pi}{2}, -\frac{\pi}{4}, 0, \frac{\pi}{4}, \frac{\pi}{2}$

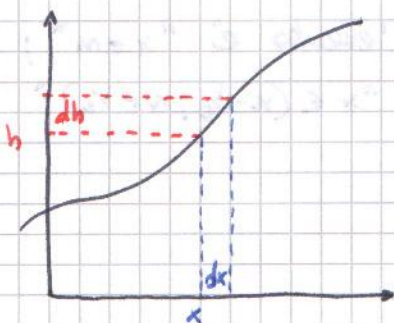
$$P(x) = \frac{1}{35} \quad \frac{3}{35} \quad \frac{6}{35} \quad \frac{10}{35} \quad \frac{15}{35}$$

$$h(x) = \cos(x) = 0 \quad 0.71 \quad 1 \quad 0.71 \quad 0$$

$$P(h(x) = h_i) = \frac{16}{35} \quad \frac{13}{35} \quad \frac{6}{35} \quad \left(\frac{13}{35}\right) \quad \left(\frac{16}{35}\right)$$

Consideriamo invece una variabile continua, ed h biettiva.

$$X \sim f(x) \quad P(x \in (x, x+dx)) = f(x) dx$$



$$P(H \in (h, h+dh)) = h(x)dh$$

Non posso fare semplicemente:

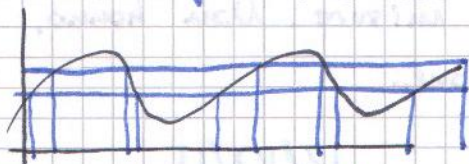
$$h(x) \rightarrow x = h^{-1}(h)$$

Devo usare la Jacobiana, in modulo, per ottenere sempre $P > 0$.

$$dh h(x) = h(x(h)) \left| \frac{dx}{dh} \right| dh$$

Questo deriva dal fatto che l'analisi matematica considera orientato il dx .

Se ho una funzione così:



$\forall h$ ho più valori di x che contribuiscono a $P(h \in (h, h+dh))$.

$$P(h_0 \in (h, h+dh)) = \sum_{x_i: h(x_i)=h_0} f(x) \left| \frac{\partial x}{\partial h} \right| dh$$

ESEMPLO

$$X \sim N(0,1) \quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$h = x^2 \Rightarrow x = \pm\sqrt{h}$$

$$g(h)dh = \frac{1}{\sqrt{2\pi}} \left(e^{-\frac{h}{2}} + e^{-\frac{h}{2}} \right)$$

$$\left| \frac{\partial x}{\partial h} \right| = \frac{1}{2\sqrt{h}}$$

$$\Rightarrow g(h)dh = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{h}} e^{-h/2} = \chi_1^2(h)$$

FUNZIONE χ^2 PER 1 GRADO DI LIBERTÀ

$$\chi_1^2(h) = \frac{1}{\Gamma(\frac{1}{2}) 2^{\frac{1}{2}}} h^{-\frac{1}{2}} e^{-h/2}$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi} \quad d = \frac{1}{2} \quad \beta = 2$$

$$= \frac{1}{\Gamma(d)\beta^d} h^{d-1} e^{-h/\beta}$$

$\Gamma(x)$ si chiama "funzione gamma".

Consideriamo la trasformazione:

$$X \sim f(x) \quad F(x) = \int_{-\infty}^x f(z) dz \quad 0 \leq F \leq 1$$

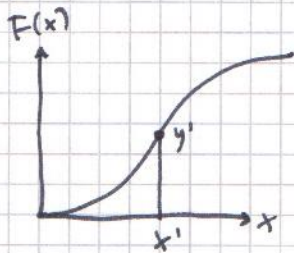
$$F(x) = P(X < x)$$

La trasformazione $y \equiv F(x)$ è in realtà molto semplice.

$$g(y) dy = f(x(y)) \left| \frac{\partial x}{\partial y} \right| dy$$

$$\frac{\partial x}{\partial y} = \frac{1}{\frac{\partial y}{\partial x}} = \frac{1}{f(x)} \quad \text{per il teorema di Leibniz}$$

$$\Rightarrow g(y) dy = \frac{f(x)}{f(x)} dy \Rightarrow g(y) \text{ è costante}$$



$$y = F(x) = P(X < x)$$

$$P(y < y') = P(X < x') = y'$$

$$g(y) = \frac{\partial P(y < y')}{\partial y} = 1$$

ESERCIZIO

$$X \sim \frac{4}{3} x^{1/3} \quad 0 \leq x \leq 1$$

- Verificare che $\frac{4}{3} x^{1/3}$ $0 \leq x \leq 1$ è una densità di probabilità ($\int = 1$)

- $y \equiv \sqrt[3]{x}$, calcolare $g(y)$

ESERCIZIO (Funzione di Pareto)

$$X \sim \frac{a}{x^{a+1}} \quad x \geq 1 \quad a > 0$$

- $y \equiv \ln x \rightarrow$ dovremmo ottenere una distribuzione esponenziale.



$$\text{Supponiamo } X \sim f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$y = a + bx$$

$$g(y) = f\left(x = \frac{y-a}{b}\right) \left|\frac{\partial x}{\partial y}\right| = f\left(\frac{y-a}{b}\right) \left|\frac{1}{b}\right|$$

$$= \frac{1}{|b|\sqrt{2\pi}} e^{-\frac{(y-a)^2}{2b^2}}$$

È sempre una distribuzione normale, ma traslata ed allungata.

$$b \equiv \sigma \Rightarrow \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-a)^2}{2\sigma^2}}$$

- Metodo corretto -

$$X \sim f(x) \quad H = g(x)$$

Dobbiamo usare le probabilità fin dall'inizio.

$$P(H \leq h) = \int_{x: g(x) \leq h} f(x) dx$$

Rifacciamo l'esempio precedente: $H \equiv a + bx$

$$P(H \leq h) = \int_{a+bx \leq h} f(x) dx = \int_{x \leq \frac{h-a}{b}, \text{ se } b > 0} f(x) dx \equiv F_H(h)$$

$$x \geq \frac{h-a}{b}, \text{ se } b < 0$$

$$g(h) = \frac{\partial F_H(h)}{\partial h} = \frac{\partial}{\partial h} \int_{-\infty}^{+\infty} f(x) dx = -\frac{1}{b} f\left(\frac{h-a}{b}\right), \text{ se } b < 0;$$

$$\text{se } b > 0 \quad g(h) = \frac{\partial}{\partial h} \int_{-\infty}^{\frac{h-a}{b}} f(x) dx = \frac{1}{b} f\left(\frac{h-a}{b}\right)$$

13/10/11

$$X \sim f(x) \quad H = g(x)$$

$$1.) \quad x = g^{-1}(H)$$

$$2.) \quad w(h) dh = f(x = g^{-1}(H)) \left|\frac{\partial x}{\partial h}\right| dh$$

$$= \sum_{x_i: g(x_i) = h} f(x_i) \left|\frac{\partial x_i}{\partial h}\right| dh$$

Ad esempio: $X \sim N(0,1)$ $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}$

$$H \equiv x^2 \quad x = \pm\sqrt{H} \Rightarrow w(h) = \frac{1}{\sqrt{2\pi h}} e^{-h/2}$$

Questa è la distribuzione di probabilità del x^2 .

Metodo matematico: $X \sim f(x)$ $h \equiv g(x)$

$$P(H \leq w) = \int_{x: g(x) \leq w} f(x) dx$$

A questo punto, $H \sim \frac{\partial P(H \leq w)}{\partial w}$

Ad esempio: $X \sim f(x)$ $g(x) \equiv a + bx$

$$P(g(x) \leq w) = \int_{x: g(x) \leq w} f(x) dx = \int_{x: a+bx \leq w} f(x) dx =$$
$$= \begin{cases} \int_{-\infty}^{\frac{w-a}{b}} f(x) dx & \text{se } b > 0 \\ \int_{\frac{w-a}{b}}^{+\infty} f(x) dx & \text{se } b < 0 \end{cases}$$

$$F_X(w) \equiv \int_{-\infty}^w f(t) dt$$

$$\Rightarrow P(g(x) \leq w) = \begin{cases} F_X\left(\frac{w-a}{b}\right) & b > 0 \\ 1 - F_X\left(\frac{w-a}{b}\right) & b < 0 \end{cases}$$

Esempio - esercizio

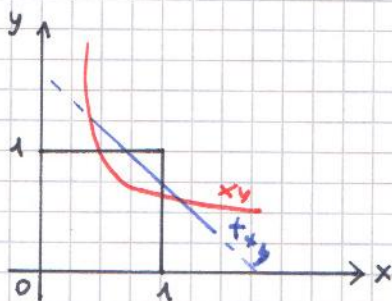
x, y iid $\sim U(0,1)$

$$z_1 \equiv x + y$$

$$z_2 \equiv xy$$

$$z_3 \equiv x/y$$

Trovare la densità di probabilità.



$P(z \leq w)$ è l'area sottesa alla curva.

Possiamo anche usare un metodo di Monte Carlo.

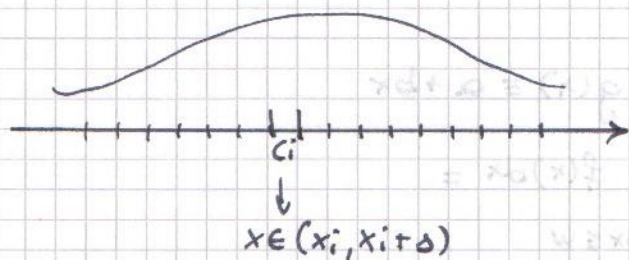
Montecarlo:

(...)
 $x = \text{rand}();$
 $y = \text{rand}();$
 $x = x + y;$
Fill histogram

restituisce un numero a caso tra 0 e 1

per questo passaggio è meglio usare Mathematica, Matlab o ROOT.

Chiamiamo E_i l'evento in cui $x \in (x_i, x_i + \delta)$



Alla fine del ciclo avremo ottenuto n_i eventi della classe c_i .

$$\frac{n_i}{N} = \text{frequenza}$$

$$= \int_{x_i}^{x_i + \Delta} f_2(x) dx \underset{\Delta \rightarrow 0}{\approx} f_2\left(x_i + \frac{\Delta}{2}\right) \Delta$$

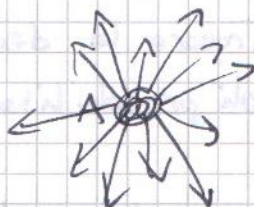
Questo metodo è molto utilizzato nella pratica.

Supponiamo che una sorgente emetta particelle in tutte le direzioni con una certa attività A , e

noi usiamo un contatore che registra un rate

$R = A \cdot \Delta\Omega$, $\Delta\Omega$ angolo solido coperto.

$$\frac{N(\text{raggi e contatore})}{N(\text{raggi emessi})} = \Delta\Omega$$



$\Delta\Omega \rightarrow$ ACCETTANZA

$$R = A \Delta\Omega$$

Esercizio

$$x, y \text{ iid } \sim N(0, 1)$$

$$z = \frac{x}{y} \quad \text{calcolo la densità di prob. di } z.$$

$$\text{Risultato: } z \sim \frac{1}{\pi} \frac{1}{1+z^2} \quad \text{funzione di Cauchy}$$

Potrebbero esserci delle singolarità. Infatti la funzione di Cauchy non ha né medio né varianza \Rightarrow usare metodo analitico.

Esercizio

x, y indipendenti

$$f(x, y) = f_x(x) f_y(y) \quad (\text{fattorizzabile})$$

$$z = x + y$$

$$\bullet f(z) = P(z \leq a) = \int_{x+y \leq a} f(x, y) dx dy =$$

$$= \iint f_x(x) f_y(y) dy dx = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{a-y} f_x(x) dx \right) f_y(y) dy =$$

$$= \int_{-\infty}^{+\infty} f_y(y) \int_{-\infty}^{a-y} f_x(x) dx dy =$$

$$= \int_{-\infty}^{+\infty} f_y(y) F_x(a-y) dy$$

$$\bullet g(a) da = \int_{-\infty}^{+\infty} f_y(y) dy \frac{\partial}{\partial x} F_x(a-y) = \int_{-\infty}^{+\infty} f_y(y) f_x(a-y) dy$$

E' ancora una volta una considerazione.

Applicazione pratica:

$$f_x(x) = \frac{1}{z} e^{-x/z} \quad (x \geq 0) \quad f_y(y) = \frac{1}{z} e^{-y/z} \quad (y \geq 0)$$

$$g(a) = \int_0^a \frac{1}{z^2} e^{-\frac{a-y}{z}} e^{-\frac{y}{z}} dy = \frac{a}{z^2} e^{-a/z}$$

Calcolare la densità di probabilità di $z = (x+y) + w$ con w ancora esponenziale.

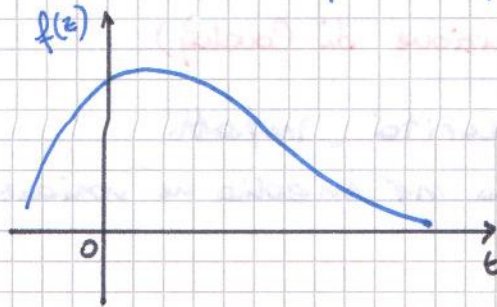
Esercizio

$$X \equiv \exp(\tau)$$

$$Y \equiv N(0, \sigma^2)$$

$$Z = X + Y$$

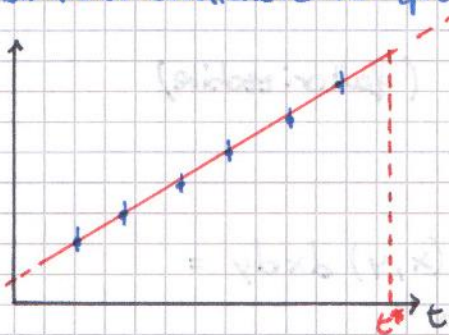
→ misura dei tempi con apparati numerosi!



Calcolare la densità di probabilità di z .

← Risultato

Supponiamo di voler controllare lo spread BTP-BUND in funzione del tempo:



Ci viene richiesta una previsione del valore ad un istante t futuro.

Facciamo un'interpolazione $y = a + bt$ e calcoliamo

il valore che assume a $t = t^*$.

Dovremo trovare \hat{a} , \hat{b} che minimizzino l'errore del fit.

$$\hat{\theta} \equiv \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$$

$$\text{Var}(\theta) = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix}$$

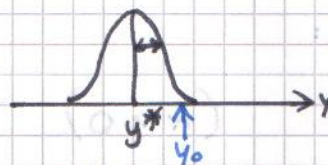
$$y^* = \hat{a} + \hat{b}t^*$$

Solitamente ogni punto è ottenuto da una media \Rightarrow una distribuzione gaussiana.

\hat{a} , \hat{b} sono una qualche funzione delle y_i : \Rightarrow verosimilmente saranno distribuite normalmente. Con che media e che varianza?

Ci attendiamo:

$$t \equiv t^*$$



$$\sigma = ?$$

Ci serve conoscere questa varianza per eventualmente confrontare il valore che abbiamo calcolato con, ad esempio, il valore massimo dello spread imposto dallo BCE (y_0).

Il test che dobbiamo fare è un test d'ipotesi.

È ragionevole pensare che y^* non assuma valori distorti.

$$V(y^*) = \text{Var}(\hat{a}) + t^{*2} \text{Var}(\hat{b}) + t^* \text{Cov}(\hat{a}, \hat{b})$$

La linea interpolante ha delle incertezze che crescono quadraticamente con t .

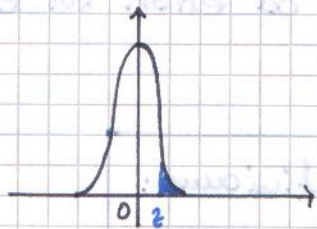
Per stabilire se y_0 e y^* sono consistenti una grandezza da considerare è

$$z = \frac{y^* - y_0}{\sqrt{\text{Var}(y^*)}}$$

Questa è una distribuzione normale riscolata:

z va paragonata con 0.

Il **p-value** è una misura di quanto sono audaci male le cose.



Un errore del 10% sulla variante può provocare un fattore 2 sulle code.

PROPAGAZIONE DEGLI ERRORI

Dobbiamo studiare come gli errori si propagano.

In questo modello, \hat{a} e \hat{b} sono correlati.

Supponiamo di imporre $\hat{a} = \hat{a} + \epsilon \Rightarrow$ il fit cambia inclinazione, per minimizzare i residui.

\hat{a} e \hat{b} sono sempre **NEGATIVAMENTE CORRELATI**:

$+\epsilon$ su \hat{a} comporta $-\delta$ su \hat{b} .

Supponiamo di avere $x_1 \dots x_n$ ed $y(x_1 \dots x_n)$.

Vogliamo una stima rapida dei valori di y sugli x_i .

Per fare questo si espande la y :

$$y(x_1, \dots, x_n) = y(\vec{\mu}) + \sum_i (x_i - \mu_i) \frac{\partial y}{\partial x_i} \Big|_{\vec{\mu}} + \dots$$

Il valore d'aspettazione è: $E(y) = y(\vec{\mu})$

$$\text{Var}(y) = E(y - E(y))^2 = \sum_{i,j} E((x_i - \mu_i) \frac{\partial y}{\partial x_i} \Big|_{\vec{\mu}} \frac{\partial y}{\partial x_j} \Big|_{\vec{\mu}} (x_j - \mu_j))$$

$$= \sum_{i,j} \text{Cov}(x_i, x_j) \frac{\partial y}{\partial x_i} \Big|_{\vec{\mu}} \frac{\partial y}{\partial x_j} \Big|_{\vec{\mu}}$$

(quando $i = j$ $\text{Cov} = \text{Var}$)

Ad esempio: se $Y \equiv X + Y$, $\text{Var}(Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$.

Consideriamo una funzione y di una singola variabile x . Possiamo scriverla come:

$$y(x) = y(\mu) + (x - \mu) \left. \frac{\partial y}{\partial x} \right|_{\mu} + \frac{1}{2} (x - \mu)^2 \left. \frac{\partial^2 y}{\partial x^2} \right|_{\mu} + \dots$$

$$E(y) = y(\mu) + \frac{1}{2} \sigma^2 \left. \frac{\partial^2 y}{\partial x^2} \right|_{\mu} \quad \sigma^2 = \text{varianza di } x$$

Gli ordini superiori al primo sono trascurabili se ϵ è piccola, rispetto a σ^2 , la variazione di y in funzione di x : bisogna vedere quanto x ed y sono correlate.

Un test d'ipotesi analizza le code di una distribuzione, quindi queste approssimazioni ai primi ordini sono pericolose e fortemente sensibili ad errori sul calcolo della varianza.

La covarianza è limitata. Se definiamo:

$$\rho \equiv \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x) \text{Var}(y)}}$$

allora ρ è compreso tra -1 e 1 .

Per semplicità di notazione definiamo:

$$S \equiv \left(\left. \frac{\partial y}{\partial x_1} \right|_{\vec{\mu}}, \dots, \left. \frac{\partial y}{\partial x_n} \right|_{\vec{\mu}} \right) \vec{\mu} \quad (\text{derivate calcolate in } \vec{\mu})$$

$$y \approx y(\mu) + S \cdot (\vec{x} - \vec{\mu})$$

$$E((y - y(\mu))^2) = E(S \cdot (\vec{x} - \vec{\mu}) (\vec{x} - \vec{\mu})^T S^T) = S \text{Cov}(x) S^T$$

Allo stesso modo, se ho

$$y_1(x_1, \dots, x_n)$$

$$y_2(x_1, \dots, x_n)$$

...

$$y_n(x_1, \dots, x_n)$$

Posso definire

$$S = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

$$\Rightarrow V(y) = S V_x S^T$$

Esercizio

$$r = \sqrt{x^2 + y^2}$$

$$\varphi = \tan^{-1}(y/x)$$

$$V_{x,y} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$$

Calcolare V_r, φ .

18/10/11

TEORIA DEL CAMPIONAMENTO

Matematici: "facciamo un campionamento da una distribuzione".

Fisici: "facciamo un esperimento n volte".

CAMPIONAMENTO: estrazione da un'urna.

Analisi del campionamento \rightarrow analisi di alcune frequenze.

Dalla scelta del campione possono derivare errori grossolani.

Supponiamo di avere un esperimento che può dare un certo numero di risultati $x_1 \dots x_m$. NB.: $x_1 \dots x_m$ sono in realtà i risultati dell'esperimento e non la variabile aleatoria.

VARIABILI ALEATORIE $\equiv X_1 \dots X_m$

Media campionaria $\bar{X} = \frac{1}{m} \sum X_m$

Varianza $V(\bar{X}) = \frac{1}{m} \sum (x_i - \mu_i)^2$ $\mu_i = E(x_i)$ $Var(x_i) = \sigma^2$

Vedremo che \bar{x} è un buon stimatore di μ . Però la prima è una variabile aleatoria, la seconda un parametro.

$$S_m = \sum X_m$$

$$\phi_{x_i}(t) = E(e^{tx_i}) \xrightarrow{t \rightarrow 0} 1 + \mu_1 t + \mu_2 \frac{t^2}{2} + \dots, \mu_m = E(x_i^m)$$

$$\begin{aligned} \phi_S(t) &= E(e^{t \sum x_i}) \rightarrow 1 + m \mu_1 t + m \mu_2 \frac{t^2}{2} + \dots \\ &= \left(1 + \mu_1 t + \mu_2 \frac{t^2}{2} + \dots \right)^m \end{aligned}$$

⇒ posso mettere in relazione i momenti di S e quelli delle variabili stesse.

$$(1 + \mu_1 t + \mu_2 \frac{t^2}{2} + \dots)^n = 1 + n\mu_1 t + n\mu_2 \frac{t^2}{2} + \binom{n}{2} (\mu_1 t)^2 + \mathcal{O}(t^3)$$

$$= 1 + n\mu_1 t + n\mu_2 \frac{t^2}{2} + \frac{n(n-1)}{2} (\mu_1 t)^2 = 1 + n\mu_1 t + \frac{t^2}{2} (\mu_2 n + \mu_1^2 n(n-1))$$

Quindi:

$$m_1 = n\mu_1 \rightarrow \bar{E}(S) = nE(X)$$

$$m_2 = \mu_2 n + \mu_1^2 n(n-1)$$

$$\text{Var}(S) = m_2 - m_1^2 \quad (\text{Var}(X) = E[(X-\mu)^2] = E(X^2) - E(X)^2)$$

$$\begin{aligned} \text{Var}(S) &= \mu_2 n + \mu_1^2 (n^2 - n) - n^2 \mu_1^2 \\ &= n(\mu_2 - \mu_1^2) \\ &= n \text{Var}(X) \end{aligned}$$

$$\begin{aligned} \bar{E}(S) &= n \bar{E}(X) \\ \text{Var}(S) &= n \text{Var}(X) \\ S &= \sum_{i=1}^n X_i \end{aligned}$$

Abbiamo usato la funzione generatrice dei momenti.

$$E(\sum X_i) = nE(X) \rightarrow E(\bar{X}) = E(X)$$

Il vantaggio è che: $\text{Var}(\bar{X}) = \text{Var}\left(\frac{\sum X_i}{n}\right) = \frac{\text{Var}(X)}{n}$

⇒ è meglio usare \bar{X} per ricavare il valore di μ : produrrà un risultato più preciso.

$$\bar{X} \rightarrow \mu \quad n \rightarrow \infty \quad \text{LEGGE DEI GRANDI NUMERI}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma_x^2}{n}\right) \quad n \rightarrow \infty$$

↑ questo vale \forall distribuzione per la quale \exists varianza.

Esempio. $X_i \text{ iid } \sim N(\mu, \sigma^2)$

Verificare che $\bar{X} \sim N(\dots)$

- $z \equiv (x - \mu) \frac{1}{\sigma} \sim N(0, 1)$
- $\phi = e^{-t^2/2}$
- $\phi(mt) = e^{-t^2 m/2} \rightarrow \text{Var} = \frac{1}{m} \dots$

Esempio: x_i iid $\sim \text{Poi}(\mu_i)$

$$\phi_p(t) = e^{\mu_i(e^t - 1)}$$

$$S = \sum x_i \Rightarrow \phi_S(t) = e^{\sum \mu_i(e^t - 1)} \rightarrow \text{Poi}\left(\sum \mu_i\right)$$

$$S \sim \frac{e^{-\lambda} \lambda^r}{r!} \quad r = 0, 1, 2, \dots \quad \frac{S}{n} = 0, \frac{1}{n}, \frac{2}{n}, \dots$$

$$P\left(\frac{S}{n} = \frac{i}{n}\right) = P(S = i) = \frac{e^{-\lambda} \lambda^i}{i!} \quad \left(\frac{S}{n} = \bar{X}\right)$$

Esercizio: caso binomiale.

In questi casi il teorema del limite centrale funziona sempre.

Consideriamo una variabile X e facciamo un cambio di variabile:

$$X \sim N(\mu, \sigma^2) \quad X \rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$Z^2 \sim \chi^2$$

Non conosciamo μ . Possiamo stimarla con \bar{X} .

Altra applicazione:

$$\text{Var}(X) = \frac{\sum (x_i - \mu)^2}{n}$$

Anche qui possiamo usare \bar{X} . Come cambiano queste quantità? \bar{X} è una variabile casuale che dipende da μ .

$$W = \sum (x_i - \bar{X})^2$$

A parte casi eccezionali, né medie né varianze sono note.

Teniamo presente che $X \sim N(0, \sigma^2)$.

$$W = \sum x_i^2 - 2 \sum x_i \bar{X} + \sum \bar{X}^2 = \sum x_i^2 - n \bar{X}^2$$

Ruotiamo le x_i con una trasformazione ortogonale.

$$h_i = \sum c_{ij} x_j \quad (\text{La norma è conservata: } \sum x_i^2 = \sum h_i^2)$$

In notazione matriciale: $H = CX$ $H^T H = X^T C^T C X \Rightarrow C^T C = I$

Condizione che voglio imporre:

$$h_1 = \frac{\sum x_i}{\sqrt{n}} = \bar{X} \sqrt{n}$$

$$\Rightarrow C = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

$$E(h_i) = E(\sum_{j=1}^m C_{ij} x_j) = \mu \sum_{j=1}^m c_{ij}$$

$$(C^T C)_{ij} = \sum_e c_{ie} c_{je} = \delta_{ij}$$

Imponiamo $i=1$:

$$= \sum_e \frac{1}{\sqrt{n}} c_{je} = 0 \quad \forall j \neq 1$$

$$E(h_i) = \begin{cases} E(h_1) = \mu/\sqrt{n} & i=1 \\ E(h_i) = 0 & i \neq 1 \end{cases}$$

Im analogo per le varianze:

$$\text{Var}(h_i) = \text{Var}(c_{ij} x_j) = \sigma^2 \sum_j c_{ij}^2$$

$$\sum_e c_{ie} c_{je} \xrightarrow{i=j} \sum_e c_{ie}^2$$

$$\Rightarrow \text{Var}(h_i) = \sigma^2$$

Anche $h_i \sim N(0, \sigma^2)$.

Tornando a w :

$$w = \sum_{i=1}^m x_i^2 - m \bar{x}^2 = \sum_{i=1}^m H_i^2 - m \bar{x}^2 = \sum_{i=2}^m H_i^2$$

I termini indipendenti tra loro diminuiscono di una unità.

$$w = \sum_{i=1}^m (x_i - \bar{x})^2 = \sum_{i=2}^m H_i^2$$

Questo ci permette di trovare uno stimatore di parametri della distribuzione. In questo caso ci interessa:

$$\begin{aligned} S^2 &= \sum_{i=1}^m (x_i - \bar{x})^2 & x_i &\sim N(0, \sigma^2) \\ &= \sum_{i=2}^m h_i^2 & h_i &\sim N(0, \sigma^2) \\ & & h_i^2 &\sim \sigma^2 \chi_{1,1}^2 \end{aligned}$$

$$\Rightarrow \sum_{i=2}^m h_i^2 \sim \sigma^2 \chi_{m-1}^2$$

Quanto vale σ^2 ? $\frac{S^2}{m-1}$ è uno stimatore di σ^2 .

Ovvero, se calcolo:

$$E\left(\frac{S^2}{m-1}\right) = \frac{1}{m-1} \sigma^2 E(\chi_{m-1}^2) = \frac{1}{m-1} \sigma^2 (m-1) = \sigma^2$$

Il suo valore d'aspettazione è σ^2 , non distorto.

N.B.: $\frac{s^2}{n-1}$ non è un numero, è una variabile distribuita come χ^2 ,
 $\mu = \bar{x}$, ma la distribuzione di \bar{x} è più stretta di un fattore $\frac{1}{\sqrt{n}}$.

Molte volte al posto della varianza usiamo la standard deviation,
che è $\sqrt{\sigma^2}$.

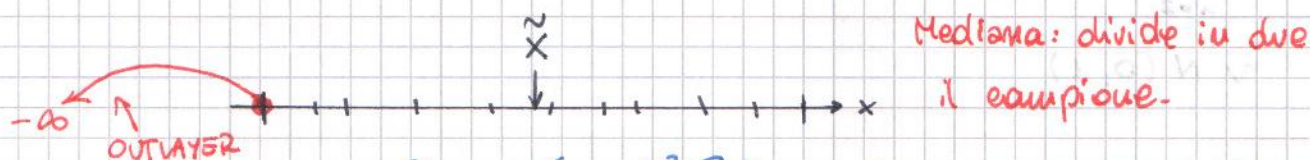
$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{s^2}{n-1}}$ è un buon stimatore per σ ? No, è
uno stimatore distorto.

$E(\hat{\sigma}) = c(n)\sigma \approx \left(1 - \frac{1}{4n} - \dots\right)\sigma$
 σ è distorto, ma è consistente! per $n \rightarrow \infty$ i bias si annullano.
Empiricamente, è meglio usare:

$$\hat{\sigma} \equiv \sqrt{\frac{s^2}{n-1.5}} \Rightarrow E(\hat{\sigma}) \approx \sigma$$

Esercizio: calcolare la covarianza $\text{Cov}(x_i - \bar{x}, \bar{x})$. Verificare
che è nulla.

Tra gli stimatori della media di una distribuzione,
abbiamo visto la media campionaria \bar{x} . Non esistono
oltre $f(x_i)$ altrettanto efficienti.



Per $n \rightarrow \infty$, $\tilde{x} \sim N\left(\mu, \frac{\sigma^2}{n} \frac{\pi}{2}\right) \Rightarrow$ la varianza è del
50% più grande rispetto a quello di \bar{x} .

Questo non è uno stimatore efficiente, ma è robusto.
È utile ad esempio in esperimenti con rumore.

Per verificare un estimatore si prende un punto
ai bordi del campione e si porta ad ∞ . La media
diverge, la mediana resta invariata. Infatti
dipende molto poco dagli outliers.

L'uso di statistiche robuste è consigliato rispetto allo
"scartamento" di un risultato outlier.

χ^2 è una variabile che deriva dalla somma di K variabili normali indipendenti.

Un'altra variabile interessante è la t di Student.

$$X \sim N(\mu, \sigma^2) \quad z = \frac{\bar{x} - \mu_0}{\sigma} \sim N(0, 1)$$



Non conosciamo μ e lo stimiamo con μ_0 .

Non conosciamo $\sigma \Rightarrow$ introduciamo:

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \leftarrow \not\sim N(\dots)$$



Una terza distribuzione molto utile è la F di Fisher.

18/10/11

S^2 è una pura variabile del χ^2 e non dipende più esplicitamente da \bar{x} .

$$S^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\sim \sum_{i=2}^n h_i^2$$

$$\sim N(0, 1)$$

$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$ ciascuno di questi oggetti è distribuito come una normale standard \Rightarrow questa somma è un χ^2 con n gradi di libertà.

$$\underbrace{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}_{\chi_n^2} = \underbrace{\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2}_{\chi_{n-1}^2} + \underbrace{\frac{n}{\sigma^2} (\bar{x} - \mu)^2}_{\chi_1^2} \quad (\text{Ho supposto } X \sim N(0, \sigma^2))$$

C'è una proprietà molto semplice:

$$\left. \begin{array}{l} X_1 \sim \chi_m^2 \\ X_2 \sim \chi_m^2 \\ X_1, X_2 \text{ indep.} \end{array} \right\} \Rightarrow X_1 + X_2 \sim \chi_{m+m}^2$$

Questo dimostra che S^2 e \bar{x} sono indipendenti.

Possiamo anche fare il processo inverso.

χ^2

Abbiamo già visto la funzione del χ^2 quando abbiamo visto i componenti.

$$X \sim N(0,1) \Rightarrow X^2 \sim \chi^2_1 \quad X^2 \equiv \mu$$

$$\mu \sim \frac{1}{\Gamma(\alpha)\beta^\alpha} \mu^{\alpha-1} e^{-\mu/\beta}, \quad \alpha \equiv \frac{1}{2} \quad \beta \equiv 2$$

$$\phi_\mu(t) = \left(\frac{1}{1-\beta t} \right)^\alpha$$

Se abbiamo k variabili normali indipendenti:

$$\sum_{i=1}^k x_i^2 = \chi^2_k$$

$$\phi_{\sum x_i^2}(t) = \left(\frac{1}{1-\beta t} \right)^{\alpha k} \quad \alpha \rightarrow k\alpha \quad \beta \rightarrow 2$$

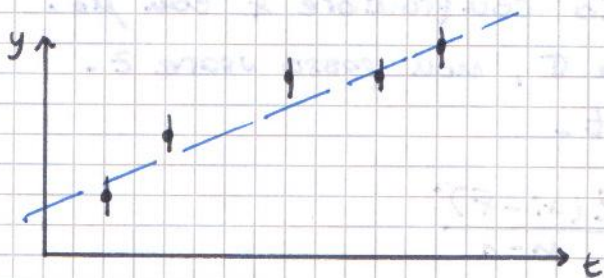
$$\mu \equiv \sum_{i=1}^k x_i^2 \sim \frac{1}{\Gamma(\alpha) 2^{\alpha k}} \mu^{\frac{k}{2}-1} e^{-\mu/2}$$

La distribuzione del χ^2 è utile per tutti i test d'ipotesi. Tornando alla dimostrazione:

$$X_1 + X_2 \sim \chi^2_{m+n}$$

Si dimostra attraverso le f. generatrici:

$$\phi_{X_1 + X_2}(t) = \left(\frac{1}{1-\beta t} \right)^{\frac{m+n}{2}} \quad \text{c.v.d.}$$



$$t_1 \quad y_1 \pm \sigma_1$$

⋮

$$t_m \quad y_m \pm \sigma_m$$

Supponiamo di fare un fit, ad esempio lineare.

È buono o no? Costruisco: $y = a + bt = f(t)$

$$\sum \left(\frac{y_i - f(t; \alpha, \beta)}{\sigma_i} \right)^2$$

Supponiamo che le t_i siano prive d'errore.

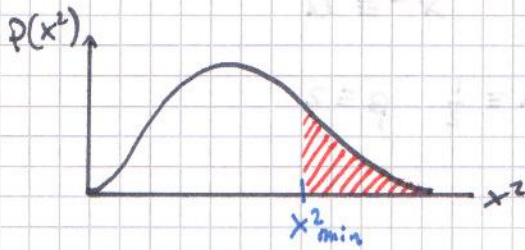
III
 χ^2

minimizzando χ^2 otengo: $\hat{a}, \hat{b}, \chi^2_{\text{min}}$

Se il fit è buono, allora χ^2 ha dei valori "ragionevoli". In fatti, ciascuno termine

$$\frac{y_i - f(x_i; a, b)}{\sigma}$$

risulta essere $\sim N$ standard
 $\Rightarrow \chi^2_{min} \sim \chi^2_{k-2}$



Supponiamo di aver trovato questo χ^2_{min} . Se rifaccio le misure tante volte, χ^2_{min} dovrebbe essere distribuito come questa curva.

Si considera la frazione di esperimenti che danno $\chi^2 > \chi^2_{min}$. Se è piccola, allora ho sbagliato qualcosa. (o siamo in un caso molto sfortunato)

Questo metodo è un test complessivo, non ci dice di preciso in che punto abbiamo sbagliato.

$$\left[P(\chi^2 > a) = \int_a^{\infty} \chi^2_{k-2} d\mu \quad a \equiv \chi^2_{misurato} \right]$$

Questo è il motivo per cui è importante la statistica del χ^2 .

t di Student

$$z = \frac{x - \mu}{\sigma}$$

Queste sono tutte distribuzioni:

collegate alle normali: $x \sim N(\mu, \sigma)$

Voglio fare dei campionamenti su una media "nota" μ_0 .

$x_1 \dots x_n \rightarrow \mu_0$ Voglio confrontare \bar{x} con μ_0 .

$$\bar{x} = \frac{\sum x_i}{n}$$

Non conosco σ , non posso usare z .

Devo usare t .

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

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

t non è distribuita come una normale standard.

$$t = \frac{\frac{x - \mu_0}{\sigma} \sqrt{n}}{\frac{\frac{s^2}{\sigma^2}}{\sigma}} \sim \frac{N}{\sqrt{\chi^2_{n-1}}}$$

Per ottenere la distribuzione di t consideriamo la funzione:

$$\Theta(N, \mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}N\mu^2} \chi^2_{n-1}(\mu) |J|$$

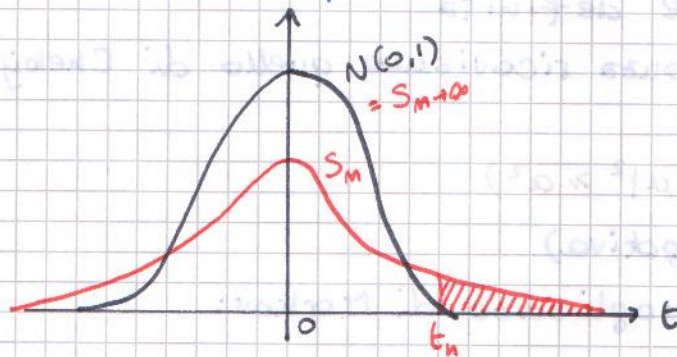
Facciamo un cambio di variabile:

$$t = \frac{N}{\sqrt{M(n-1)}} \quad Z \equiv M$$

Il risultato finale è:

$$t \sim c \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}} \Rightarrow t \sim S_{M-1}$$

Che è una distribuzione di Cauchy, tendente alla normale standard per $M \rightarrow \infty$.



Questo è un tipico esempio di come si uso il test di Student. Un altro esempio è il confronto tra due distribuzioni normali:

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_x^2 + \sigma_y^2}} \quad (\text{è una variabile correttamente normalizzata})$$

Analisi della Varianza

Se abbiamo molti campionamenti e vogliamo capire se appartengono tutti alla stessa distribuzione utilizziamo questo test.

$$X_i = N(0, \sigma_1^2) \quad S_1^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$Y_i = N(0, \sigma_2^2) \quad S_2^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$$

Se i campioni vengono dalla stessa popolazione avranno la stessa varianza. $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$\frac{S_1^2}{S_2^2} = \frac{S_1^2 / \sigma^2 \cdot m}{S_2^2 / \sigma^2 \cdot n} = \frac{\chi^2_m}{\chi^2_n} = f \text{ di Fisher}$$

Se due popolazioni sono indipendenti \Rightarrow i χ^2 sono indipendenti.

$$f \sim c' \frac{f^{m/2 - 1}}{(mf + n)^{\frac{m+n}{2}}}$$

20/10/11

Supponiamo di avere $X \sim f(x)$ $X > 0$ Calcoliamo l'integrale $E(x) = \int_0^{\infty} x f(x) dx = \int_0^a x f(x) dx + \int_a^{\infty} x f(x) dx$

$$\Rightarrow \int_0^{\infty} x f(x) dx \geq a \int_a^{\infty} f(x) dx$$

Disuguaglianza di Markov

$$P(x \geq a) \leq \frac{1}{a} E(x)$$

• $x > 0$

• La media deve essere definita

Da questa disuguaglianza ricaviamo quello di Chebyshev:

$$\mu = E(x)$$

$$P(|x - \mu| \geq a) = P(|x - \mu|^2 \geq a^2)$$

(x può anche essere negativa)

Applichiamo la disuguaglianza di Markov:

$$P(|x - \mu|^2 \geq a^2) \leq \frac{E(|x - \mu|^2)}{a^2} = \frac{\sigma^2}{a^2}$$

$$a = k\sigma$$

Disuguaglianza di Chebyshev

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

Questa è una delle più forti disuguaglianze indipendenti dalla $f(x)$. Da esso segue il teorema dei grandi numeri. Vediamo come si usano queste disuguaglianze.

Supponiamo che una fabbrica di auto abbia in media una produzione di 50 auto/settimana. Qual è la probabilità che questo numero superi 75?

$$P(x \geq 75) \leq 50/75 = 0.7 \quad \text{Non dice molto.}$$

Supponiamo di sapere che la varianza di questa produzione sia 25 \Rightarrow possiamo usare Chebyshev, con un piccolo trucco.

$$P(x \geq 75) \leq P(|x - 50| \geq 25) \leq \frac{25}{25^2} = 4\% \quad (\sigma^2 = 25)$$

Esercizio. $Y: E(Y^n) = \mu_n$

$$P(|Y| \geq t^n \sqrt{\mu_n}) \leq \frac{1}{t^2} \quad \text{Dimostrare.}$$

Esempio: la contabilità di un supermercato è approssimata all'unità di € più vicino. Qual è l'errore che si commette, a fine giornata?

$$X \sim U(-\frac{1}{2}, \frac{1}{2}) \quad m \quad E(X) = 0 \quad \text{Var}(X) = \frac{1}{12}$$

$$S = \sum_{i=1}^m X_i$$

$$P(|\sum X_i| > 500) \leq \frac{m \frac{1}{12}}{500^2}$$

Per un numero tipico di transazioni, $P \leq 3\%$.

Esempio: polling prima di una votazione.

Indicatore: $\begin{cases} 1 & \text{se candidato A} \\ 0 & \text{se candidato B} \end{cases} \quad P(A) \equiv p$

(questo indicatore è chiaramente una binomiale).

Quante persone devo interrogare per avere una precisione su p del 20%?

$I_1 \dots I_m$ iid $\Rightarrow S = \frac{\sum I_i}{m}$ è un estimatore di p ($= \hat{p}$)

$\text{Var}(S) = \frac{p(1-p)}{m}$ Voglio una precisione al 90%:

$$P(|\hat{p} - p| > 0.2) \leq \frac{p(1-p)}{m(0.2)^2} \leq 1 - 90\% = 0.1$$

Non conosco p , ma so che $0 \leq p \leq 1 \Rightarrow p(1-p) \leq \frac{1}{4}$

\Rightarrow trovo un upper limit.

$$\frac{1}{4m(0.2)^2} \leq 0.1 \Rightarrow \sim 65 \text{ persone}$$

Esercizio: A vince al 90%. La mio ditta di polling annuncerà la vittoria di A se $p > 50\%$. Qual è la probabilità che A vinca al 90% se $p = 0.6$?

Se la distribuzione ammette varianza, da Chebyshev discende la legge dei grandi numeri.

$$K\sigma \equiv \varepsilon \quad P(|x - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Applichiamolo alla media campionaria:

$$\bar{x} = \frac{\sum x_i}{m} \quad \text{Var}(\bar{x}) = \frac{\text{Var}(x)}{m} = \frac{\sigma^2}{m}$$

$$P(|\bar{x} - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 m}$$

Legge dei grandi numeri in forma debole

$\bar{x} \xrightarrow{p} \mu$, \bar{x} tende a μ in probabilità

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| > \varepsilon) = 0$$

Legge dei grandi numeri in forma forte

$$P\left(\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| > \varepsilon\right) = 0$$

La differenza è che " $\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| > \varepsilon$ " è a tutti gli effetti un evento; \bar{X}_n è la media campionaria su n elementi.



Un modo comodo per presentare questi numeri è la **distribuzione empirica**:

$$F_n(y) = \frac{\# x_i < y}{n}$$

La sua versione statistica è: $F(y) = \int_{-\infty}^y f(x) dx$ $x_i \sim f(x)$
 Tramite la legge dei grandi numeri:

si dimostra che $F_n \rightarrow F$. F_n è una variabile binomiale:

$$F_n \sim B(F(y), n)$$

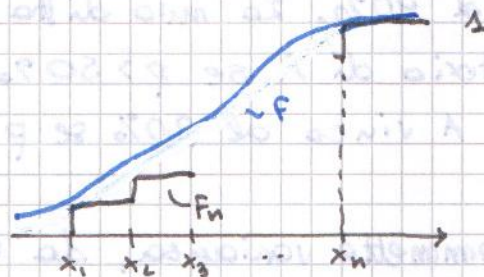
$$E(F_n(x)) = F(x) \quad \text{Var}(F_n(x)) = \frac{F(x)[1-F(x)]}{n}$$

Applicando la disuguaglianza

di Chebyshev:

$$P(|F_n(x) - F(x)| > \varepsilon) \leq \frac{F(x)(1-F(x))}{n\varepsilon^2} \leq \frac{1}{4} \frac{1}{\varepsilon^2} \frac{1}{n}$$

$\Rightarrow F_n$ tende ad F puntualmente ($\forall x \in \mathbb{R}$) (teorema di Cantelli)



Supponiamo di avere:

$$x_1, \dots, x_n \quad f(x, a) \quad E(x)(a) \quad E(x) = h(a)$$

$$\bar{x} = \frac{\sum x_i}{n}$$

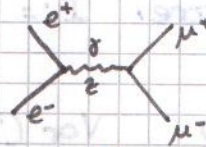
$$a \equiv h^{-1}(E(x)) \quad \bar{x} \xrightarrow{P} E(x)$$

$\Rightarrow \hat{a} = h^{-1}(\bar{x})$ tende in probabilità al valore vero:

STIMA DEI PARAMETRI

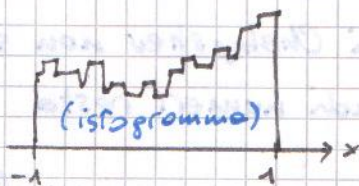
$\hat{a} \xrightarrow{n \rightarrow \infty} a$ È un metodo estremamente semplice.

Esempio: $e^+e^- \longrightarrow \mu^+\mu^-$



L'interferenza tra il diagramma col γ e quello con lo Z produce un termine di violazione di parità nella sezione d'urto: $\frac{d\sigma}{d\cos\theta} \propto (1 + \cos^2\theta + \alpha \cos\theta)$

Questo esperimento serve per trovare α . Si possono ad esempio prendere $\sim 10^4$ eventi e ricostruire $\frac{d\sigma}{d\cos\theta}$ per frequenza: $\cos\theta = x$



m_i : # ev. $x \in [x_i, x_i + dx]$

$$m_i = N \int_{x_i}^{x_i + dx} f(x, \alpha) dx$$

$$\sum_i \frac{(m_i - n_i)^2}{m_i} = \chi^2 \dots$$

(Ricostruzione della distribuzione di probabilità di $\cos\theta$)

C'è anche un metodo più semplice:

primo di tutto trovare la normalizzazione

$$k: \int_{-1}^{+1} k(1+x^2+\alpha x) dx = 1$$

$$\bar{x}(x) = \int_{-1}^{+1} x k(1+x^2+\alpha x) dx = \frac{\alpha}{4} = \mu \quad \text{Questo è la funzione da invertire.}$$

$$\alpha = 4\mu \quad \hat{\alpha} = \bar{x} \cdot 4$$

Questo metodo non è storicamente rigoroso, ma è consistente.

L'altro metodo consiste nel confrontare frequenza empirica con probabilità:

$$\frac{\# x \in C}{N} \longleftrightarrow \int_{x \in C} f(x) dx$$

Si cerca di far corrispondere la distribuzione con l'istogramma variando α (metodo dei minimi quadrati): questo produrrà un altro estimatore $\hat{\alpha}'$. Si dimostra che $\hat{\alpha}$ ed $\hat{\alpha}'$ convergono.

Qual è il problema su $\hat{\alpha}$? L'errore relativo è 16 volte quello su \bar{x} .

$$x \quad P(x \in C) = \int_C f(x) dx \equiv p$$

Definiamo un indicatore: $I_i = \begin{cases} 1 & x \in C \\ 0 & x \notin C \end{cases}$

$$E(I) = p$$

$$\bar{I} = \frac{\sum I_i}{n} \quad E(\bar{I}) = p \quad \text{Var}(\bar{I}) = \frac{p(1-p)}{n}$$

$$P(|\bar{I} - p| > \epsilon) < \frac{p(1-p)}{n \epsilon^2}$$

Abbiamo ritrovato la stessa espressione con Kolmogorov (con numero campionamenti $\rightarrow \infty$).

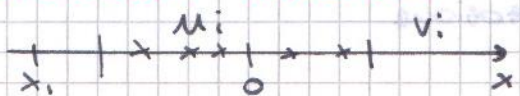
Questo giustifica l'utilizzo di istogrammi e montecarlo per campionare distribuzioni difficili da calcolare analiticamente.

Se la varianza non è finita la legge di Chebyshev non si può più applicare, ma la legge dei grandi numeri resta vera.

$$P\left(\left|\frac{\sum x_i}{n} - \mu\right| > \epsilon\right) \rightarrow 0 \quad (\text{non sappiamo con quale velocità}).$$

Metodo delle "troncazioni":

$$|x_i| \leq \delta_n \rightarrow$$



$$P(|\sum u_i| \geq \frac{n\epsilon}{2}) \rightarrow 0$$

$$P(|\sum v_i| \geq \frac{n\epsilon}{2}) \rightarrow 0$$

$$P(|\sum u_i| \geq \frac{n\epsilon}{2}) \leq \frac{n \text{Var}(\sum u_i)}{\epsilon^2 n^2} \quad (\text{se } \exists \text{Var}(\sum u_i))$$

Possiamo giocare sui due parametri n e δ .

$$E(|x_j|) = \sum_j |x_j| f(x_j) \equiv a$$

$$E(|u_i|^2) = \sum_{|x_i| < \delta_n} x_i^2 f(x_i) \leq \delta_n \sum_{|x_i| < \delta_n} |x_i| f(x_i) \leq \delta_n a$$

$$E(u_i) = \sum_{|x_i| \leq \delta_n} x_i f(x_i) \xrightarrow{n \rightarrow \infty} E(x_i) = 0 \quad (\text{ho supposto che tutte queste distribuzioni fossero centrate in 0})$$

$$\text{Var}(u_i) = \delta_n a \quad \text{Var}(\sum u_i) = n^2 \delta_n a$$

$$\Rightarrow P(|\sum u_i| \geq \frac{1}{2} \epsilon n) \leq \frac{\frac{1}{2} \epsilon n}{\frac{1}{2} \epsilon n} a \delta \xrightarrow{\delta \rightarrow 0} 0$$

l'altro termine:

$$P(|\sum v_i| \geq \frac{1}{2} n \epsilon) \leq P(|\sum v_i| \neq 0) \leq \sum_i P(v_i \neq 0) = n P(v_i \neq 0)$$

$$= n \sum_{|x_i| > \delta_n} f(x_i) \leq n \sum_{|x_i| > \delta_n} \frac{x_i}{n \delta} f(x_i) \quad (n \delta \leq x_i)$$

$$= \sum_{|x_i| > \delta_n} |x_i| f(x_i) \frac{1}{\delta} \xrightarrow{n \rightarrow \infty} 0$$

La legge dei grandi numeri è valida anche per distribuzioni che non hanno varianza.

Esercizio al pc.: Compilare una distribuzione di Cauchy e verificare la legge dei grandi numeri.

Catena di Markov = serie di stati ognuno dei quali dipende solo dal precedente.

Si crede che la legge dei grandi numeri valga anche in questo caso (NB: finora abbiamo esplicitamente richiesto l'indipendenza degli eventi.)

25/10/11

- $X_1 \dots X_n$ indipendenti, con distribuzioni $F_1(x_1) \dots F_n(x_n)$
- $E(x_i) = \mu_i < \infty$
- $\text{Var}(x_i) = \sigma_i^2 < \infty$ almeno uno > 0
- $S_n \equiv \sum_i x_i$ $\sigma_n \equiv \sqrt{\sum \sigma_i^2}$

Teorema del limite centrale di Lindberg

allora

$$\frac{S_n - E(S_n)}{\sigma_n} \xrightarrow[n \rightarrow \infty]{\text{in distrib.}} N(0, 1)$$

Lo dimostriamo per X_i iid: $E(x_i) = \mu$, $\text{Var}(x_i) = \sigma^2$, $\mu, \sigma^2 < \infty$

$$\frac{\sum x_i / n - \mu}{\sigma / \sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{dist.}} N(0, 1)$$

$$\phi_x(t) = 1 + m_1 t + m_2 \frac{t^2}{2} + \mathcal{O}(t^3)$$

$$z_i \equiv \frac{x_i - \mu}{\sigma \sqrt{n}}$$

Calcoliamo $\phi_{x_i - \mu}$:

$$\begin{aligned} \phi_{x_i - \mu}(t) &= 1 - E(x_i - \mu) + \frac{t^2}{2} E((x_i - \mu)^2) + \dots \\ &= 1 - 0 + \frac{t^2}{2} \sigma^2 + \mathcal{O}(t^3) \end{aligned}$$

Per il denominatore:

$$\frac{\phi_{x_i - \mu}}{\sigma \sqrt{n}} = 1 + \frac{t^2}{2} \frac{1}{n \sigma^2} \sigma^2 + \dots \phi_{w_i}(t)$$

$$\phi_{\sum_{i=1}^n w_i}(t) = \left(1 + \frac{t^2}{2n} + \dots\right)^n = e^{t^2/2}$$

Quindi: $\sum w_i \sim N(0, 1)$

$$\sum_i w_i = \sum \frac{x_i - \mu}{\sigma \sqrt{n}} = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$$

$$E(\bar{x}) = \mu \quad \text{Var}(\bar{x}) = \sigma^2/n$$

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

altro forma del teorema
del limite centrale

Questo ci assicura, ad esempio, che gli stimatori tendono asintoticamente al valore vero.

Esempio: transazioni approssimate all'unità di €.

$$x_i \sim U\left(-\frac{1}{2}, \frac{1}{2}\right) \quad E(x_i) = 0 \quad \text{Var}(x_i) = \frac{1}{12}$$

N transazioni

$$P(|\sum x_i| \geq a) \leq \frac{N}{12} \frac{1}{a^2} \quad (\text{Chebisher})$$

Possiamo invece usare il TLC:

$$\frac{\sum x_i}{N} \sim N\left(0, \frac{1}{12N}\right)$$

$$\text{Var}(\sum x_i) = \frac{N}{12} \quad \text{Var}\left(\frac{\sum x_i}{N}\right) = \frac{1}{12N}$$

$$P\left(\left|\frac{\sum x_i}{N}\right| \geq \frac{A}{N}\right) = 2 \int_{A/N}^{\infty} N\left(0, \frac{1}{12N}\right) dz$$

Per $N \equiv 100$ ed $A = 5€$ abbiamo:

$$\text{Chebisher: } P \leq 33\%$$

$$\text{TLC: } P = 8.3 \cdot 10^{-2}$$

La disuguaglianza di Chebisher è la più forte che possiamo usare in questi casi.

Soppontiamo di avere un orologio con un errore di $\pm \frac{1}{2}$ minuto per giorno.

Speriamo: 5 minuti d'errore all'anno?

$$P\left(\left|\sum_{i=1}^{365} x_i\right| < 5\right) = P(-5 \leq \sum x_i \leq 5) =$$

$$= P\left(-\frac{5}{n} \leq \frac{\sum x_i}{n} \leq \frac{5}{n}\right) \quad n \text{ è abbastanza grande, la distribuzione tende ad una } N \text{ standard.}$$

Si ricava:

$$P \approx \int_{-5/n}^{5/n} N(\bar{\mu}, \bar{\sigma}^2) dx = 69\%$$

esercizio: fare i calcoli

Esercizio In $1 \mu s$, l'emissione di k elettroni da una superficie calda è distribuita con:

$$P(k) = \frac{e^{-\mu} \mu^k}{k!} \quad \mu = 0.1 \mu s^{-1}$$

Calcolare il numero di elettroni emessi in $1 s \rightarrow 99500 \div 101000$
 $P(n \text{ e in } 1 \text{ secondo compreso tra } 99500 \text{ e } 101000) = \dots$

Esempio La probabilità che una CPU commetta un errore è 10^{-14} . Il pc effettua 1 operazione ogni $100 ns$.

$P(>100 \text{ errori in } 3 \text{ anni}) = ?$

secondi = $3 \cdot \pi \cdot 10^7 \text{ sec}$

$\sim 10^{15}$ operazioni

La probabilità d'errore è binomiale con p estremamente piccola \Rightarrow calcolo difficile.

Usando invece il TLC, $P \geq 94\%$.

Esempio voglio misurare la distanza d di una stella - Le misure x_i (iid) mi danno una sigma $\sigma = 2$.

$$E(x_i) = d \quad \text{Var}(x_i) = 4 \quad \text{Var}\left(\frac{\sum x_i}{n}\right) = \frac{4}{n}$$

$$\frac{\sum x_i}{n} \sim N\left(d, \frac{4}{n}\right)$$

Combiniamo variabile: $z = \frac{\bar{x} - d}{2/\sqrt{n}} \sim N(0, 1)$

$$P(|z| < 0.5) \geq 95\%$$

$$P(-.5 \leq z \leq .5) \geq 95\%$$

$$P\left(-\frac{\sqrt{n}}{4} \leq \bar{x} - d \leq \frac{\sqrt{n}}{4}\right) = 95\% \Rightarrow n \approx 65 \text{ osservazioni necessarie}$$

Non tutte le variabili aleatorie si prestano all'uso del TLC - La distribuzione di Cauchy (= distribuzione **Lorentziana**) non ha varianza finita - la massa dello z^0 segue questa distribuzione. In questo caso anche la media $\bar{x} \sim$ Cauchy. Questo si ricava utilizzando la f. generatrice delle probabilità anziché quella dei momenti.

TEORIA DEI TEST STATISTICI

Carl Pearson \rightarrow test del χ^2 (1900)

Fisher (anni '20 - '30) \rightarrow test di significato \rightarrow reude popolare il test di Pearson.

Anni '50 → **teoria dei test**. (Niemann, Pearson figlio)

(≡ teoria delle)

Metodo di Fisher: si paragonano dati sperimentali ad una "**ipotesi nulla**", sulla quale viene costruita la distribuzione dei dati, e vengono cercate discrepanze.

Uso il p-value come strumento.

Metodo Niemann - Pearson: no ipotesi nulla, occorrono 2 ipotesi tra cui scegliere.

In entrambi i casi, l'ipotesi nulla è quella che vogliamo falsificare.

~ ~ ~ **Test di significato** ~ ~ ~

• cerco di provare che l'ipotesi nulla è incompatibile con i dati.

Problema della signora e del tè. Fisher propone di preparare 4 e 4 tazze di tè, che vengono randomizzate. La signora deve classificarle.

H_0 ≡ la signora è incapace di distinguere (≡ ipotesi nulla)

4 tazze A, 4 tazze B ⇒ il modo con cui posso combinare 8 tazze in 2 gruppi da 4 è $\binom{8}{4} = 70$.

L'evento in cui tutte le tazze sono nella partizione corretta è solo 1 dei 70.

$$P(4 \text{ corrette} | H_0) = 1/70$$

Fisher: se $P(\dots | H_0) > 5\%$, allora H_0 va bene, altrimenti: devo cambiare il test.

$$P(3 \text{ corrette, 1 sbagliata} | H_0) = \binom{4}{3} \binom{4}{1} \cdot \frac{1}{70} = 16/70$$

$$P(2 \text{ corrette, 2 sbagliate} | H_0) = \binom{4}{2} \binom{4}{2} \cdot \frac{1}{70} = 36/70$$

$P(4 \text{ corr.} | H_0)$ è piccolissima ⇒ forse H_0 è falsa.

$P(3, 1 | H_0)$ è più realistico. Fisher: devo considerare anche effetti periferici → $P(3, 1 | H_0) + P(4 | H_0)$.

Supponiamo, a rovescio:

$$P(4, 0) = 16/70$$

ovvero supponiamo che le

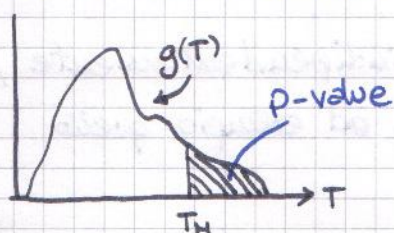
$$P(3, 1) = 1/70$$

probabilità siano sbagliate.

62 ~~Supponiamo, a rovescio:~~ (...)

$$p\text{-value} = P(X \geq X_{\text{mis.}} | H_0)$$

- definiamo un'ipotesi H_0 che sia statisticamente soddisfacente.
↳ H_0 è supposto vero
- definisco, data la variabile casuale T , la sua distribuzione $g(T | H_0)$
- faccio le misure, $T_N = T(x_1, \dots, x_n)$



Se quest'area è molto piccolo, allora l'esperimento è critico dal punto di vista significativo.

NB.: è stato fatto un solo esperimento, che ha dato il valore T_N . Gli altri punti della curva di frequenza sono puramente teorici, costruiti data H_0 .

Frequentisti: $T(x_1, \dots, x_n)$ è una variabile casuale che posso ri-sampolare n volte, e posso ricostruire la curva di frequenza.

Per i Bayesiani questo non è possibile.

NB.: \nexists conferma di H_0 tramite il p-value. Si può solo confutare nel caso $p < 5\%$.

Nel caso dello signora e del tè \nexists ipotesi alternativa: se si "conferma" H_0 non si conferma nulla.

Questo metodo è applicabile ad ogni esperimento: ciò causa molti dubbi, ma è comunemente accettato.

26/10/11

Fisher - "Design of experiments" - teoria dei test statistici - p-value

- Si parte da una ipotesi che vogliamo testare:
"ipotesi nulla" H_0
- Assumiamo vero H_0 e costruiamo una distrib. massimamente descrittiva dell'esperimento:
 $T = T(x_1, \dots, x_n)$ "statistica"
- Calcoliamo $g(T | H_0)$

- Si definisce **p-value** $P(T > |t_{crit}| | H_0)$, con t_{crit} valore misurato di T nell'esperimento.

Questa disgiunzione logico può indurre a rigettare H_0 , se il p-value è molto piccolo - Fisher: non si deve rigettare H_0 solo in base al p-value!

I fisici invece usano il p-value in maniera secca, ignorando il consiglio di Fisher:

$$p\text{-value} < \text{soglia} \Rightarrow \text{rigetto } H_0$$

Il più delle volte però la scelta è operata indipendentemente, usando altri metodi accanto al p-value, ed esempio quello dei minimi.

→ Vedi esperimento quark liberi su disperse.

Prima di annunciarne una scoperta, c'è bisogno non solo di un esperimento ad alta **significanza**, ma anche di un esperimento che sia convincente agli occhi della comunità scientifica.

Esempio: violazione parità osservata negli anni '30 su elettroni provenienti da decadimento β (quindi **polarizzati!**) - Risultati non accettati.

Esempio: scoperta positrone, '30. L'articolo non conteneva alcuna formula e fu comunque subito accettato.

N.B.: frequentisticamente, \nexists modo per calcolare $P(H_0 | \text{dati})$

In quanto H_0 non è una variabile aleatoria, bensì un fatto.

Devo saggiamente direse H_0 e trovare quella che non sia sconfermata dai dati, al più -

Da un punto di vista Bayesiano, potremmo trovare:

$$P(H_0 | \text{dati}) = \frac{P(\text{dati} | H_0) P(H_0)}{P(\text{dati})}$$

che tuttavia dipende fortemente dal prior.

\exists tuttavia un metodo (Neymann-Pearson) che consente un calcolo simile: confrontiamo H_0 con una diversa ipotesi H_1 .

$$P(H_1 | \text{dati}) = \frac{P(\text{dati} | H_1) P(H_1)}{P(\text{dati})}$$

$$\frac{P(H_0 | \text{dati})}{P(H_1 | \text{dati})} = \frac{P(\text{dati} | H_0)}{P(\text{dati} | H_1)} \frac{P(H_0)}{P(H_1)}$$

fattore di Bayes

Il rapporto $P(H_0)/P(H_1)$ è più verosimilmente calcolabile. Il test decisionale viene effettuato così:

$$\begin{cases} \frac{P(H_0 | \text{dati})}{P(H_1 | \text{dati})} > 1 & \Rightarrow \text{viene usata } H_0 \\ \frac{P(H_0 | \text{dati})}{P(H_1 | \text{dati})} < 1 & \Rightarrow \text{viene usata } H_1 \end{cases}$$

M.B.: le ipotesi nulle devono essere statisticamente solide, ovvero devono consentire previsioni precise.

Es.: Lanciamo 100 monete \Rightarrow 40 teste

$$H_0: "p = 1/2"$$

$$\hat{p} = \frac{\sum I_T}{n} \quad I_T = \begin{cases} 1 & \text{testa} \\ 0 & \text{croce} \end{cases}$$

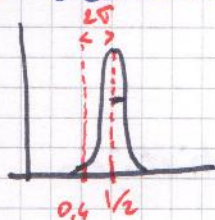
$$\hat{p} = 0.4 \quad E(\hat{p}) = p \quad E(\hat{p} | H_0) = 0.5$$

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{100} \quad \text{Var}(\hat{p} | H_0) = \frac{1}{400}$$

$$\text{Stand. Dev.}(\hat{p} | H_0) = \sqrt{\frac{1}{400}} = \frac{1}{20} = 0.05$$

Senza scomodare il TLC, una media è distribuita come una normale.

Posso misurare la distanza tra $E(\hat{p})$ e p in unità di standard deviation:



$$d = \frac{0.1}{0.05} = 2 \Rightarrow \text{questo risultato è significativo}$$

Il p-value è la somma delle code a destra e a sinistra, in quanto H_0 non è una "diseguaglianza".

Per il TLC: $\hat{p} \sim N(p_0, \frac{p_0(1-p_0)}{n})$ $p_0 \equiv E(\hat{p} | H_0)$

$$z \equiv \frac{p_0 - \hat{p}}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Il test è fatto con: $P(|z| > z_{\text{mis}}) \approx 2\%$
("test a due code")

M.B.: possiamo anche testare z^2 , $z^2 \sim \chi_1^2$.

27/10/11

Supponiamo di avere k insiemi di variabili: $k: x_i^i \quad i=1 \dots k$ (indipendenti)

$$\bar{x}_i \sim N(\mu_i, \frac{\sigma^2}{n})$$

Ipotesi: $H_0: \mu_1 = \dots = \mu_k$

Definisco $z_i = \frac{\bar{x}_i - \mu_i}{\sigma} \sqrt{n}$

Per testare H_0 dovrei fare k test paralleli. È delicato: se ne fallisce anche solo 1, H_0 è da rigettare.

Definisco invece:

$$u = z_1^2 + \dots + z_k^2 \sim \chi_k^2$$

Possiamo testare solo u , ciò risponde alla medesima domanda.

Non stiamo testando k variabili indipendentemente.

Se $k-1$ variabili sono consistenti con H_0 e 1 no, il p-value non è tale da far rigettare H_0 .

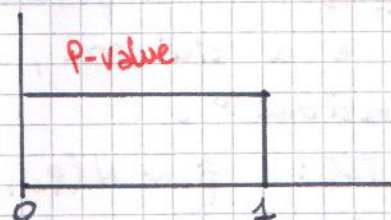
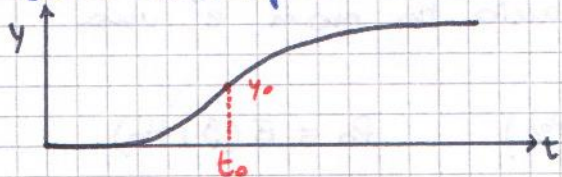
Se testiamo le k variabili abbiamo più informazioni, ma il modo di combinare k p-value in modo da dare lo stesso risposta che darebbe il test su u è complicato.

$$p\text{-value} = P(T > t_m | H_0)$$

t_m aleatoria \Rightarrow p-value è una variabile aleatoria.

$$y = 1 - p\text{-v} \Rightarrow y = P(T < t_m | H_0)$$

Questa è la definizione della distribuzione statistica di T , ed è del tipo:



$$P(y \leq y_0) = P(t \leq t_0) = y_0$$

\Rightarrow Se H_0 è vera, il p-value è distribuito uniformemente; se non lo è è fortemente spostato verso 0.

Il fatto che p-value sia distribuito uniformemente ci permette di combinare più test allo Fisher.

Il test di **Neyman-Pearson** confronta due ipotesi alternative, ed elabora l'errore che si commette rigettando "per sbaglio" l'una o l'altro.

$$X \sim N(\mu, \sigma^2)$$

$$Z \equiv \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n} \quad \mu_0 = \mu | H_0$$

Ad esempio: H_0 : L'Higgs non esiste
 Ma se sono fortemente convinto che l'Higgs esista, per qualche sua piccola manifestazione,
 $\mu_0 \rightarrow \mu_0 + \delta$

Il test statistico che vorrei fare è

$$P(Z \geq z_\alpha | \delta)$$

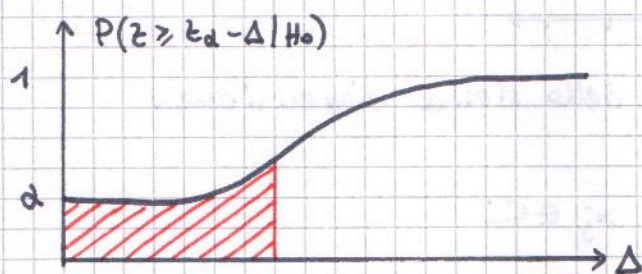
$$Z' \equiv \frac{\bar{x} - (\mu_0 + \delta)}{\sigma} \sqrt{n} = Z - \Delta \quad \Delta = \frac{\delta \sqrt{n}}{\sigma}$$

Definiamo un "livello d'attenzione" α tale che:

$$Z > z_\alpha \Rightarrow p\text{-value} < \alpha$$

$$2 \int_{z_\alpha}^{\infty} N(0,1) dz = \alpha$$

Il mio test è quindi: $P(Z \geq z_\alpha - \Delta | H_0)$ che è una funzione di Δ :



Questa funzione si dice **potenza del test**.

Testarla significa coprire l'errore che si commette supponendo $\Delta = 0$.

Esempio: la frazione delle stelle di classe A nei cluster aperti deve essere 0.1.

$$n = 95 \text{ osservazioni} \Rightarrow N_A = 5$$

Queste informazioni non sono sufficienti per fare un test aleatorio: non sappiamo come è stata fatta la misura:

- osservazione di 95 open clusters \sim binomiale
- osservazione di abbastanza open cluster per trovare $N_A = 5 \sim$ binomiale negativa (Pascal)

Nel primo caso p -value $\approx 12\%$, nel secondo $\approx 3\%$.

Nella statistica frequentistica occorre avere sempre ben chiaro come è fatto il campione: le code della distribuzione sono determinate non empiricamente, ma teoricamente basandosi sulla distribuzione statistica del campione.

Quando si fa un esperimento, come si fa a sapere qual è il campione di riferimento?

Ad esempio: ricerca dello z^0 . Potrei fare la ricerca a tempo fisso, oppure a luminosità totale fissa.

Primo test statistico proposto da Pearson

$$x_1 \dots x_N \quad H_0: X \sim f(x)$$

Questo è essenzialmente un test del χ^2 .

Può essere fatto con una distribuzione empirica:

$$F_m(x) = \frac{\# x_i < x}{m}$$

Divido l'intervallo in k classi:



Non c'è bisogno che siano della stessa dimensione.

Definiamo:

$$I_j^i = 1 \text{ se } x_j \in C_i$$

$$r_i = \sum_{j=1}^m I_j^i \quad (\text{occupazione di } C_i)$$

Ognuna delle r_i è una variabile distribuita binomialmente.

$$P(x \in C_i) = \int_{C_i} f(x) dx = p_i$$

$$E(r_i) = m p_i \quad V(r_i) = m p_i (1 - p_i)$$

Per $m \rightarrow \infty$ posso assumere che:

$$r_i \sim N(m p_i, m p_i)$$

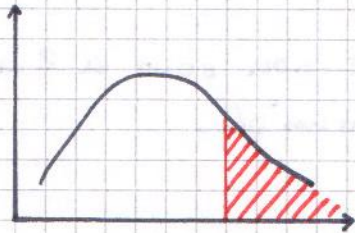
Definisco:

$$z_i = \frac{r_i - mp_i}{\sqrt{mp_i}} \sim N(0, 1) \quad (m > 5)$$

Risulta quindi:

$$\sum z_i^2 \sim \chi_{k-1}^2 \quad (\text{constraint: } r_k = m - \sum r_i)$$

Posso fare il test del p-value su questa distribuzione.



C'è qualcosa che non funziona.

Cioè se r_i è distribuito binomialmente; però l'insieme dei k r_i non è binomiale, è **multinomiale**.

Come è uno storico multinomiale?

Supponiamo di avere ^{estrazioni di} m palline di 2 colori e di estrarre: 1 1 2 1 2 1 2 2 2 1 2 1 1 ...

Ci interessa sapere m_1, m_2

$$P_1 + P_2 = 1$$

$$m_1 + m_2 = m$$

$$P(m_1 | m, P_1) = \binom{m}{m_1} P_1^{m_1} (1 - P_1)^{m - m_1}$$

Supponiamo invece di avere 3 tipi di palline.

$$P_1 + P_2 + P_3 = 1$$

$$m_1 + m_2 + m_3 = m$$

$$\binom{m}{m_1} \binom{m - m_1}{m_2} = \frac{m!}{m_1! m_2! (m - m_1 - m_2)!} = \frac{m!}{m_1! m_2! m_3!}$$

$$P(m_1, m_2, m_3 | P_1, P_2, P_3, m) = \binom{m}{m_1, m_2, m_3} P_1^{m_1} P_2^{m_2} P_3^{m_3} \quad \text{tri-nomiale}$$

$$P(m_1, \dots, m_k | m, \vec{p}) = \frac{m!}{m_1! \dots m_k!} p_1^{m_1} \dots p_k^{m_k}$$

$$E(m_i) = m p_i$$

$$E(m_i \cdot m_j) = \sum_{i \neq j} \frac{m!}{\dots} m_i m_j p_i^{m_i} p_j^{m_j} (1 - p_i - p_j)^{m - m_i - m_j}$$

$$= m(m-1) \left[\sum \frac{(m-2)!}{(m_i-1)!(m_j-1)!} p_i^{m_i-1} p_j^{m_j-1} (\dots)^{m - (m_i-1) - (m_j-1) - 2} \right]$$

$$= p_i p_j m(m-1) \quad \cdot p_i p_j$$

$$\text{Cov}(m_i, m_j) = E(m_i m_j) - E(m_i) E(m_j) = -m p_i p_j$$

NB.: le variabili di una multinomiale non sono a 2 a 2 correlate al 100% come le binomiali.

Averemo scritto:

$$\chi^2 = \frac{(\bar{r} - m\bar{p})^2}{m\bar{p}} = (\bar{r} - m\bar{p})^T V_r^{-1} (\bar{r} - m\bar{p})$$

$$V = \begin{pmatrix} m_1 p_1 & & & 0 \\ & m_2 p_2 & & \\ & & \dots & \\ 0 & & & m_k p_k \end{pmatrix} \Rightarrow V^{-1} = \text{Diag} \left(\frac{1}{m p_i} \right)$$

Questo è il caso BINOMIALE -

Nel caso multinomiale:

$$V_r = m \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & -p_1 p_3 & \dots & -p_1 p_k \\ -p_1 p_2 & p_2(1-p_2) & & & \\ & & \dots & & \\ & & & & \\ & & & & \end{pmatrix}$$

Questa è una matrice simmetrica che non ha rango completo \Rightarrow determinante = 0 \Rightarrow \nexists inverso.

Ma esiste un inverso generalizzato $\sim \chi_N^2$ con $N = \text{rango}(V)$.

$$\Delta \equiv \text{Diag}(p_1 \dots p_k)$$

$$\pi \equiv \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} \Rightarrow V_r = m(\Delta - \pi \pi^T)$$

Eliminiamo l'ultima riga: $k \rightarrow k-1$

e calcoliamo il determinante di questa nuova matrice.

$$\det(C_0) = p_1 p_2 \dots p_{k-1} \begin{vmatrix} 1-p_1-p_k \\ \vdots \\ 1-p_k \\ \vdots \\ p_i \\ \vdots \\ \end{vmatrix} (\dots)$$

Risulta:

$$(C_0)_{ij} = P_i(1-P_i)\delta_{ij} - P_iP_j(1-\delta_{ij})$$

$$(C_0)^{-1}_{ij} = \frac{1}{P_j}\delta_{ij} + \frac{1}{P_k}$$

Esercizio: verificare $C_0^{-1}C_0 = I$

Posso quindi calcolare:

$$\begin{aligned} X^2 &= \frac{1}{n} \sum_{i,j=1}^{k-1} (r_i - mp_i) C_0^{-1}_{ij} (r_j - mp_j) \\ &= \sum_{i=1}^k \frac{(r_i - mp_i)^2}{mp_i} \end{aligned}$$

Che è lo stesso risultato che avevamo già calcolato col metodo sbagliato.

I matematici usano le matrici inverse generalizzate:

$$A^{-1}A = I \longrightarrow A A_g A = A$$

$$C_- \equiv \text{Diag} \left\{ \frac{1}{mp_i} \right\}$$

$$X^2 = (\bar{r} - n\bar{p})^T C_- (\bar{r} - n\bar{p})$$

Teorema

$$\begin{array}{ll} X \text{ multinomiale} & X \sim N(0, C) \\ X^T C_- X \sim \chi_N^2 & N = \text{rank } C \end{array}$$

08/11/2011

Z-test

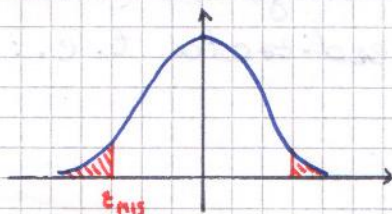
H_0 da testare con n dati: $t(x_1 \dots x_n)$

$g(t | H_0)$

p-value = $P(t > t_{mis} | H_0)$

Per il test costruisco $X \sim N(\mu, \sigma^2)$

$H_0: \mu = \mu_0 \quad Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$ se H_0 è vero



$$\text{p-value} = 2 \int_{|z_{mis}|}^{\infty} N(0, 1) dz$$

test di Pearson

Organizzare i punti sperimentali in classi non necessariamente delle stesse dimensioni. Se le classi sono k :

$$x_1 \dots x_m \xrightarrow{\text{discretizzazione}} r_1 \dots r_k$$

A differenza delle r_i , le x_i sono tutte indipendenti.

$$P_i = (x \in C_i) = \int_{C_i} f(x) dx$$

$$\sum r_i = n$$

$$r_1 \dots r_k$$

$$m_1 \dots m_k \quad m_i \equiv m p_i$$

Per n molto grande, $z_i \equiv \frac{r_i - m p_i}{\sqrt{m p_i}} \sim N(0, 1)$

$$u \equiv \sum z_i^2 \sim \chi_{k-1}^2 \quad (\text{c'è un vincolo, cioè } k-1 \text{ gradi di libertà})$$

Questa derivazione è sbagliata: $z_i \sim$ ~~binomiale~~,
 $z_i \sim$ multinomiale.

Avremo introdotto una matrice di covarianza:

$$P(m_1, \dots, m_k) = \underbrace{p_1^{m_1} \dots p_k^{m_k}}_{\text{parte probabilistica}} \underbrace{\frac{n!}{m_1! \dots m_k!}}_{\text{coefficiente } \neq \text{binomiale}}$$

$$\sum P_i = 1$$

$$\sum m_i = n$$

MATRICE DI COVARIANZA

$$C_{ij} = -m p_i p_j \quad \text{se } i \neq j \\ = m p_i (1 - p_i) \quad \text{se } i = j$$

supponendo tutte le r_i indipendenti tra di loro la matrice è diagonale.

$$\text{Definisco: } x_i = r_i - m p_i$$

$$\sigma_i = \sqrt{m p_i}$$

$$V \equiv \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \dots & \\ 0 & & & \sigma_k^2 \end{pmatrix}$$

$u = \sum \left(\frac{r_i - m p_i}{\sqrt{m p_i}} \right)^2 = X^T V^{-1} X$ Questa definizione non va bene nel caso multinomiale. Dobbiamo invece usare:

$$u \equiv X^T C_{NN}^{-1} X$$

Il problema è che questa matrice è singolare (determinante nullo). Si definisce un'inversa generalizzata t. c.:

$$~~C^{-1} C = 1~~$$

$$C \tilde{C} C = C$$

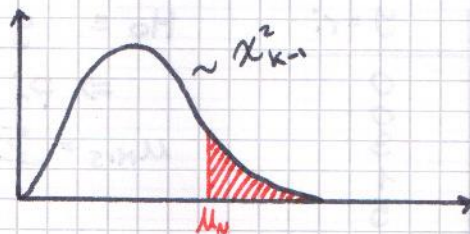
La nostra matrice di covarianza non ha rango completo
 \Rightarrow elimino una riga ed una colonna, tenendo a mente
 che $\sum n_i = n$ e che $n_k = n - \sum_{i=1}^{k-1} n_i$

$$u = \sum_{i,j=1}^{k-1} x_i x_j C_{ij}^{-1} = \sum_{i=1}^k \left(\frac{r_i - n p_i}{\sqrt{n p_i}} \right)^2$$

Adesso u è ben definita ed è lo stesso risultato
 con una multinomiale e con una binomiale.
 La nostra ipotesi H_0 ha come statistica u :

$$u = \sum z_i^2 \sim \chi_{k-1}^2$$

$$= \sum \frac{(r_i - n p_i)^2}{n p_i}$$



Questo test è:

- **asintotico** (non funziona per n piccoli)
- **non parametrico** (non dipende da $f(x)$).

Il test z è parametrico. È un po' più potente.
 In genere i test non parametrici sono meno potenti.

È più conveniente utilizzare binning stretti, per
 vedere la distribuzione con più precisione. Ma
 bin troppo stretti \Rightarrow bassa statistica.

Empiricamente si usa:

$$\begin{array}{l} \text{numero di eventi: } m \\ \text{numero di classi: } k \end{array} > 5 \text{ numero di eventi per bin}$$

Per definire le classi definisco $F(x) = \int_{-a}^x f(z) dz$
 (distribuzione di probabilità),
 dopodiché dividiamo l'asse y in k classi uguali.
 In questo modo le classi hanno per definizione lo
 stesso contenuto di probabilità \Rightarrow in principio nel
 sampling avranno un contenuto uguale di eventi.



$$F_i = \frac{i}{k}$$

$$x = F^{-1}(F_i)$$

Nel 99% dei casi la distrib. che vogliamo studiare dipende da un parametro θ da stimare dai dati $\rightarrow \hat{\theta}$.

- ogni volta che assumiamo un $\hat{\theta} = \hat{\theta}(x_1 \dots x_n)$ a partire dai dati, il x^i corrispondente si ottiene di un grado di libertà.

$$H_0 = X \sim f(x, \theta = \hat{\theta})$$

Esercizio: suddivisione asse x:

x	y = r _i
0 - 0.2	0
0.2 - 0.4	0
0.4 - 0.6	3
0.6 - 0.8	1
0.8 - 1.0	5

$$H_0 \equiv y = \text{cost}$$

$$\Rightarrow p_i | H_0 = 0.2 \quad \forall i$$

$$\chi^2_{MIS} = \sum \frac{(r_i - np_i)^2}{np_i}$$

Fare m simulazioni:

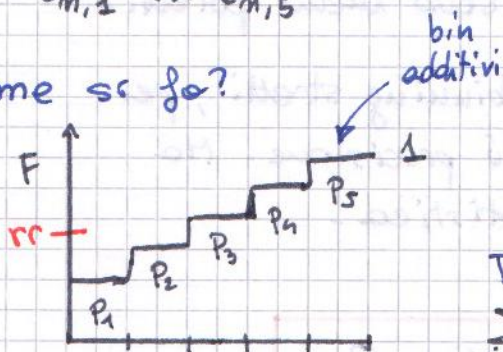
$$t_{1,1} \dots t_{1,5}$$

$$t_{2,1} \dots t_{2,5}$$

$$\dots$$

$$m \rightarrow \infty$$

Come si fa?



$$\equiv \text{rdm}() ?$$

$$rr := \text{random}(); |rr| \leq 1$$

Trovare in che intervallo casca rr.

$$\sum_{i=1}^j P_i \leq rr < \sum_{i=1}^{j+1} P_i$$

$$\chi^2_{MC} = \sum_{i=1}^5 \frac{(t_{i,m} - np_i)^2}{np_i}$$

\Rightarrow confrontare con χ^2_{MIS}

Questo test è parametrico.

$$rr := |\text{random}| < 1$$

\hookrightarrow classe = #i

$$r_i ++$$

k volte

$$r_i \equiv t_i$$

$$\{t_i\} \rightarrow \chi^2_{MC}$$

conto quante volte $\chi^2_{MC} > \chi^2_{MIS}$

l volte

$$\Rightarrow p\text{-value} = \frac{\# \text{ iterazioni}}{l}$$

N.B.: in questo caso $k=9$ (dimensione sample = $0+0+3+1+5$)

Esempio y_i intensità della linea α dello stella i -esima

H_0 : y_i indipendente dalle distanze del cluster

$$H_0: y_i = \text{cost} = k$$

$$y_i \sim N(k, \sigma_i^2)$$

Supponiamo di aver fatto m misure (y_i, σ_i) .

$$u \equiv \sum \frac{(y_i - k)^2}{\sigma_i^2} \sim \chi_m^2$$

$$P\text{-value} = P(u > u_m | H_0)$$

Supponiamo che $p\text{-v} \sim 10^{-6} \Rightarrow$ quando i dati sperimentali:



l'andamento costante è giusto ma la costante è sbagliata \Rightarrow

faccio un altro test d'ipotesi supponendo

$$H_0 \equiv y_i = \text{cost} = c(y_i)$$

Stimo \hat{c} ed metodo del χ^2 :

$$u = \sum \left(\frac{y_i - c}{\sigma_i} \right)^2 \quad \hat{c} \leftarrow \text{minimizza } u$$

questo \hat{c} non minimizzerebbe il $p\text{-v}$.

$$\frac{\partial u}{\partial c} = -2 \sum \frac{y_i - c}{\sigma_i^2} \equiv 0 \Rightarrow \hat{c} = \frac{\sum y_i / \sigma_i^2}{\sum 1 / \sigma_i^2} \quad \text{MEDIA PESATA}$$

\hat{c} è una variabile aleatoria tanto quanto le y_i .

\Rightarrow non posso dire che $\frac{y_i - c}{\sigma_i} \sim N(0, 1)$.

C'è però un teorema:

$$u_{\min} \equiv \sum \left(\frac{y_i - \hat{c}}{\sigma_i} \right)^2 \sim \chi_{m-1}^2$$

TEOREMA

Possiamo riscrivere l'ipotesi ridotta come:

$$H_0 \equiv E(y_i) = c$$

$$(y_i \sim N(c, \sigma_i^2))$$

$$\hat{c} = \frac{\sum y_i / \sigma_i^2}{\sum 1 / \sigma_i^2} \equiv \sum w_i y_i \quad w_i = \frac{1 / \sigma_i^2}{\sum 1 / \sigma_i^2}$$

$$E(\hat{c}) = \sum w_i E(y_i) = \sum w_i \cdot c = c$$

$$\text{Var}(\hat{c}) = \sum w_i^2 \text{Var}(y_i) = \sum w_i^2 \sigma_i^2 = \frac{1}{\sum 1 / \sigma_i^2}$$

Supponiamo di avere:

$$X \sim N(\mu, \sigma^2) \quad \sigma^2 \text{ non nota}$$

$$X_1 \dots X_m$$

$$H_0 \equiv \mu = \mu_0$$

Posso costruire: $\bar{x} = \frac{\sum x_i}{n}$

ma non posso definire $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ perché non conosco σ .

$$s^2 \equiv \sum_{i=1}^m (x_i - \bar{x})^2 \sim \sigma^2 \chi_{m-1}^2$$

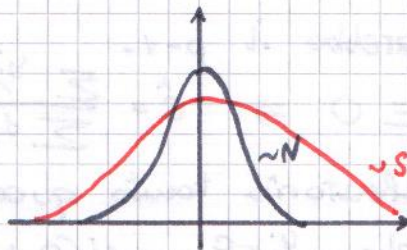
$E(s^2) = \sigma^2 (m-1) \Rightarrow \frac{s^2}{m-1}$ è uno stimatore per σ .

$$\frac{s^2}{m-1} \equiv s_x^2$$

$$t \equiv \frac{\bar{x} - \mu_0}{s_x/\sqrt{n}}$$

Questa non è più una $N(0,1)$ ma è una Student con $n-1$ d.o.f.:

$$t \sim S_{m-1}(x)$$



Per $m=2$ la $S(x)$ è una Cauchy.

In questo caso il test è a due code:

$$p\text{-v} = 2 \int_{|t_H|}^{\infty} S_{m-1}(t) dt$$

TEST DI OMOGENEITÀ

Supponiamo di avere:

$$X \sim N(\mu_1, \sigma_1^2)$$

$$Y \sim N(\mu_2, \sigma_2^2)$$

$$H_0 \equiv \mu_1 = \mu_2 \wedge \sigma_1^2 = \sigma_2^2$$

Ovvero le due variabili sono estratte dallo stesso campione.

n misure di X \bar{x}

m misure di Y \bar{y}

σ_1^2, σ_2^2 non conosciute.

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{m-1}$$

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2}{m-1}$$

$$t = \frac{\bar{x} - \bar{y}}{\text{variata}}$$

cosa metto qui? s_x^2, s_y^2 , o una combinazione?

$$S^2 \equiv f S_x^2 + (1-f) S_y^2$$

f va scelto in modo da minimizzare S^2 .

N.B.:

$$\frac{S_x^2 (m-1)}{\sigma^2} \sim \chi_{m-1}^2$$

$$\frac{S_y^2 (m-1)}{\sigma^2} \sim \chi_{m-1}^2$$

Sono indipendenti \Rightarrow la somma $\bar{e} \sim \chi_{\text{somma}}^2$ d.o.f. :

$$\frac{S_x^2 (m-1)}{\sigma^2} + \frac{S_y^2 (m-1)}{\sigma^2} \sim \chi_{m+m-2}^2$$

$$\hat{\sigma}^2 \equiv \frac{S_x^2 (m-1) + S_y^2 (m-1)}{m+m-2}$$

Posso quindi definire: $t = \frac{\bar{x} - \bar{y}}{\hat{\sigma} \sqrt{\frac{1}{m} + \frac{1}{m}}}$

Ma se non siamo sicuri che le varianze di x e y debbano essere =, ci conviene ridefinire

$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_x^2}{m} + \frac{S_y^2}{m}}}$ (così facendo, rinunciamo al test di omogeneità e confrontiamo soltanto l'uguaglianza delle medie) -

Anche i test di Student sono combinabili per testare

n campioni aleatori: **analisi della varianza-covarianza** -

Allora la funzione test diventa: $t = \frac{\bar{x} - \bar{y}}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{1}{m}}}$ \Leftrightarrow μ e σ sono uguali per x e y

Se medie e σ non sono uguali:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \quad E(s_x^2) = \sigma_x^2$$

$$E\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right) = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$$

\rightarrow Questa è std dei termini che stanno al numeratore

Questo test è importante perché prescinde dalla conoscenza della varianza della popolazione.

Il problema che si pone è se ho parecchie distribuzioni e voglio vedere se i valori sono compatibili tra di loro.

Per vedere se diverse misure sono compatibili tra loro si può fare a 2 a 2, ma così non si vede la loro compatibilità stiamo solo combinando varie distribuzioni tra loro. Per fare ciò si usa l'analisi della varianza-covarianza, e un'analisi della compatibilità delle medie tra di loro.

09 novembre 2011

$$x_i \quad i = 1 \dots n$$

$$y_j \quad j = 1 \dots m$$

Possiamo confrontare tra loro due campioni così:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} \sim S(t) \quad g^2 = \frac{s_x^2(n-1) + s_y^2(m-1)}{n+m-2}$$

Questo vale solo se le varianze sono uguali, se pensiamo che siano diverse allora non va bene.

Vogliamo ora paragonare delle cose, per vedere se dai diversi siti di produzione sono compatibili tra loro, fare il test 2 a 2 non va bene.

Le misure che abbiamo sono:

$$x_{1,1} \quad x_{1,2} \quad \dots \quad x_{1,u_1}$$

$$x_{2,1} \quad x_{2,2} \quad \dots \quad x_{2,u_2}$$

$$\vdots$$

$$x_{p,1} \quad \dots \quad x_{p,u_p}$$

Conosciamo che $X \sim N(\mu, \sigma^2)$

I parametri sono uguali

Questo test si chiama test di Hotelling

H_0 : omogeneità

$$SS = \sum_{i,j} (x_{i,j} - \bar{x})^2 \quad \text{essendo} \quad \bar{x} = \frac{\sum_{i,j} x_{i,j}}{n} \quad n = \sum u_i$$

Anche questo test è stato inventato da Fisher.

$$= \sum_{i,j} (x_{i,j} - \bar{x}_j + \bar{x}_j - \bar{x})^2 =$$

$$= \underbrace{\sum_{i,j} (x_{i,j} - \bar{x}_j)^2}_{SS_0} + \sum_{j=0} (\bar{x}_j - \bar{x})^2$$

$$\sum_{j=0} (\bar{x}_j - \bar{x})^2 \cdot n_j$$

$$SS \sim \chi_{n-1}^2 \cdot \sigma^2$$

$$SS_B = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2 \sim \chi^2_{n-p}$$

χ^2_{n-p} \hookrightarrow Dopo aver sommato su j

SS_W : χ^2_{p-1} ma direi così non va bene, si deve ricordare il teorema di Fisher.

Teorema di Fisher

Variabili X distribuite come $N(0, 1)$, (questo è il nostro caso se H_0 fosse vera).

$$Z = AX \quad A = X^T X = Z^T Z = B + C$$

Supponiamo che B sia una forma quadratica $B = \sum z_i^2$, anche C è fatta così ma con un numero diverso di dof

$$B \sim \chi^2_p \quad C \sim \chi^2_{n-p}$$

Con questo teorema si risolve il nostro problema, allora anche SS_W una variabile indipendente. $n-p$ gradi sono attribuiti a SS_B , restanti $p-1$ sono attribuiti a SS_W .

A questo punto il test è completato, si può scrivere:

$$s_B^2 = \frac{SS_B}{n-p}$$

Se ora gli s^2 sono uguali:

$$s_W^2 = \frac{SS_W}{p-1}$$

$$f = \frac{SS_W}{SS_B} \sim F_{p-1, n-p}$$

è una variabile di Fisher

p -value = $\int_{F_0}^{\infty} F_{p-1, n-p}(f) df$ e allora possiamo decidere se rigettare o no H_0

Vedere l'esempio sulla velocità della luce di Michelson.

Allora non possiamo usare il test ipotesi perché ci dà una risposta secca.

possiamo fare 2 tipi di plot: uno con il valore ottenuto e la varianza, e un altro che dà la mediana (plot scade e baft).

Quello che invece viene fatto in fisica: $F_n(x) = \frac{\sum x_i < x}{n}$ e graficare questa $F_n(x)$.

Fare quello che si fa quando abbiamo molte variabili o una generalizzazione del t-student e serve per fare quel confronto che in realtà faremo 2 o 2 che poi non sappiamo come mettere insieme i risultati.

Esempio

$$p=2 \quad n_1 = n_2 = m \quad n = 2m$$

$$w = \sum n_j (\bar{x}_j - \bar{x})^2 = m [(\bar{x}_1 - \bar{x})^2 + (\bar{x}_2 - \bar{x})^2] = \frac{m}{2} (x_1 - x_2)^2$$

In questo caso non conosciamo le varianze, al più le abbiamo, ma possiamo fare il test ugualmente.

$$SS_B = \frac{1}{n-2} \sum (x_{ij} - \bar{x}_j)^2 = \frac{1}{n-2} \sum \sum (x_{ij} - \bar{x}_j)^2 =$$

$$= \frac{m-1}{n-2} (S_1^2 + S_2^2) = \frac{S_1^2 + S_2^2}{2}$$

$$F = \frac{(\bar{x}_1 - \bar{x}_2)^2}{\frac{S_1^2}{m} + \frac{S_2^2}{m}} = t^2$$

Questo test student generalizzato usa i dati

• $X \sim f(x)$ $F(x) = \int_{-\infty}^x f(t) dt$

I risultati saranno a casaccia e quindi l'array che misuro sarà diverso da quello ordinato.

$$F_n(x) = \frac{\#X_i < x}{n}$$

→



$$\hookrightarrow \begin{cases} 0 & \text{se } x < x_1 \\ 1 & \text{se } x > x_n \end{cases}$$

$F_n \rightarrow F(x)$ per n grande

$n F_n(x) \sim \text{Bin}(p=F(x), n)$ e' una binomiale

Usando la disuguaglianza di Chebyshev:

$$P(|F_n(x) - F(x)| \geq k\sigma) \leq \frac{1}{k^2}$$

oppure

$$P(|F_n(x) - F(x)| \geq a) \leq \frac{\sigma^2}{a^2}$$

Si deve usare σ^2 della $F_n = \sigma^2 = \frac{F(x)(1-F(x))}{n}$

$$P(\dots) \leq \frac{1}{a^2} \frac{F(x)(1-F(x))}{n}$$

⇒ Questo ci dice che F_n converge alla probabilità $F(x)$ quando n e' grande

Si deve allora trovare una statistica che misuri lo scarto tra le 2 distribuzioni, ci sono 2 modi per fare questo e lo vedremo domani.

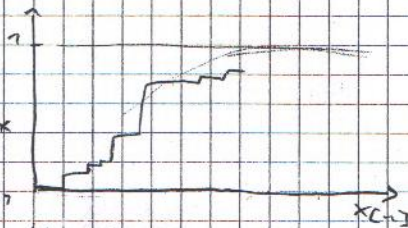
Il test di Pearson ammette ipotesi composte.

10 novembre 2011

$$X \sim f(x) \quad F(x) = \int_{-\infty}^x f(t) dt$$

x_1, \dots, x_n

$$F_n(x) = \begin{cases} 0 & \text{se } x < x_{(1)} \\ \frac{i}{n} & \text{se } x_i \leq x < x_{(i+1)} \\ 1 & \text{se } x > x_{(n)} \end{cases}$$



Qui x_i fa un salto pari a $\frac{1}{n}$ fino ad arrivare a 1

Tutto questo se H_0 e' vera

Il primo test prende il nome di ω^2 test ed e' definito come:

$$\omega^2 = \int_0^1 (F_n(x) - F(x))^2 dF =$$

comincio a spezzare l'integrale in n pezzi.

$$= \int_0^{x_{(1)}} F(x)^2 dF + \sum_{i=1}^{n-1} \int_{x_{(i)}}^{x_{(i+1)}} (F_n(x) - F(x))^2 dF + \int_{x_{(n)}}^1 (F_n(x) - F(x))^2 dF =$$

ogni pezzo e' costante

$$= \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^n \left(F(x_{(i)}) - \frac{i-1}{2n} \right)^2$$

$$t = n\omega^2$$

10/11/11

Distribuzione empirica a partire dal campionamento

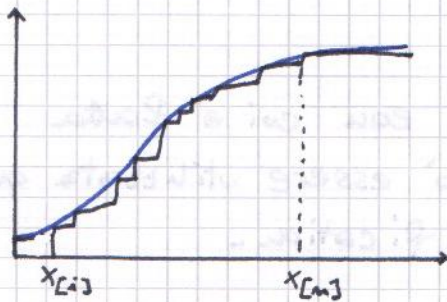
$x_1 \dots x_m$:

$x \sim f(x)$

$$F(x) = \int_{-\infty}^x f(t) dt$$

$$F_m(x) = \begin{cases} 0 & \text{se } x < x_{[1]} \\ i/m & \text{se } x_{[i]} \leq x < x_{[i+1]} \\ 1 & \text{se } x > x_{[m]} \end{cases}$$

$i = \# x_j \leq x$



Possiamo stimare la distanza tra queste distribuzioni.

ω^2 test

$$\begin{aligned} \omega^2 &\equiv \int_0^1 [F_m(x) - F(x)]^2 dF = \int_0^1 F(x)^2 dF + \sum_{i=1}^{m-1} \int_{x_{[i]}}^{x_{[i+1]}} (F_m(x) - F(x))^2 dF + \\ &\quad + \int_{x_{[m]}}^1 (F_m(x) - F(x))^2 dF = \\ &= \frac{1}{12m^2} + \frac{1}{m} \sum_{i=1}^m \left[F(x_{[i]}) - \frac{2i-1}{2m} \right]^2 \end{aligned}$$

$F(x_{[i]})$ sono i campionamenti, m lo conosciamo
 \Rightarrow possiamo calcolare ω^2 .

Calcoliamo la distribuzione asintotica ($m \geq 10$)
 di $t = m\omega^2$.

I valori critici di questo statistico sono:

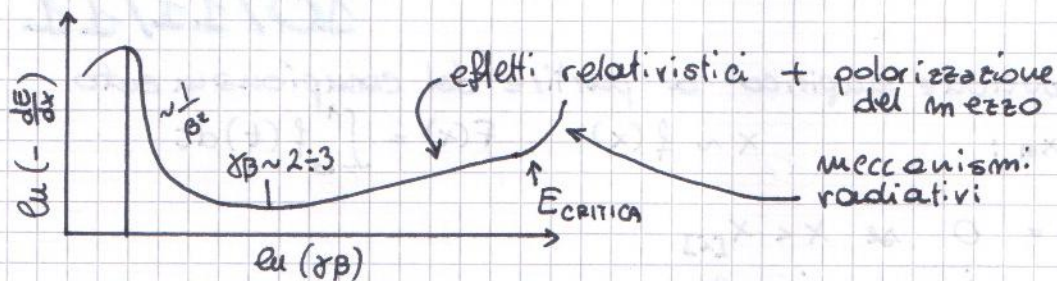
$$\alpha = 0.05 \quad t = 0.461 \quad (t \geq 0.461 \Rightarrow p.v. \leq 0.05)$$

$$\alpha = 0.01 \quad t = 0.743$$

Questo schema si usa per i test di massa in fisica delle alte energie.

È difficile misurare direttamente la massa di una particella.

Quando una particella passa attraverso un volume sensibile, ionizza il mezzo \Rightarrow quantità di luce emessa $\propto \Delta x$ della particella. È possibile riconoscere una particella dalla sua perdita d'energia specifica.



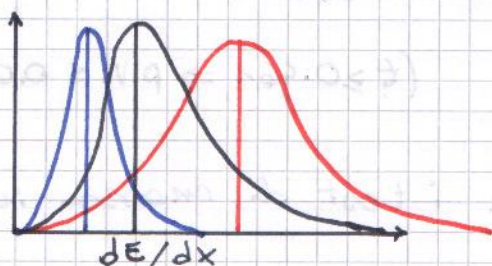
Nelle regione iniziale $\Delta E/dx$ dipende solo da β^{-2} .

$$-\frac{dE}{dx} = \frac{A}{\beta^2} + B = A \left(\frac{m}{p}\right)^2 + B$$

A e B dipendono dal materiale con cui è fatta la misura. $-\frac{dE}{dx} \propto m^2 \Rightarrow$ può essere utilizzata questa misura per fare particelle identificazioni.

Qual è la distribuzione di $-\frac{dE}{dx}$ a (m, p) fissi? La teoria inizialmente elaborata da Landau non si adatta bene ai dati sperimentali.

Ci saranno tanti piccoli urti a bassissima energia che non disturberanno la traiettoria della particella. Il TLC suggerisce una gaussiana. Ma è un teor. matematico che non tiene conto della fisica: ci sono anche effetti drastici, come l'estrazione di un elettrone in grado di ionizzare da un atomo. Questi raggi δ sono poco probabili ma hanno grandi effetti (è la coda destra delle distribuzioni) \Rightarrow gaussiane vs Landau.



Supponiamo di avere un rivelatore fatto a stati, ognuno con il suo $-\frac{dE}{dx}|_i$, i quali vanno confrontati con la distrib. di dE/dx . (La distrib. cambia media e forma al variare della particella). In questo caso ad esempio: $H_0 \equiv$ la particella è un π .

Il test w_2 va leggermente modificato per adattarlo a questo tipo di problema \rightarrow **test di Smirnov modificato**.

Questo test non è bivariato.

Il test più moderno e più utilizzato è il **test di Kolmogorov-Smirnov**, che utilizza come statistica non la dispersione delle $F_n(x)$ risp. ad $F(x)$, bensì un solo punto:

$$d_n = \sup_{-\infty < x < +\infty} |F_n(x) - F(x)|$$

la statistica è $n \cdot d_n$.

Come nel caso precedente, la distrib. non è scrivibile in termini finiti, ma solo a livello asintotico - Dobbiamo trovare dei valori critici.

Tutti questi test sono test su ipotesi semplici. In molti casi utilizziamo però questi test con valori non noti, stimati dal campione di riferimento. In questo caso va ancora bene fare il test, ma i livelli di significanza subiscono forti variazioni.

Vediamo come si calcolano in pratica questi d_n . Supponiamo che $x_1 \dots x_n$ sia il campionamento ordinato. Avindoli: $F_n(x_i) = i/n$

Definiamo: $y_i \equiv F(x_i)$

Se l'ipotesi H_0 è vera, le y_i sono distribuite uniformemente (per lo stesso teorema col quale abbiamo calcolato la distribuzione del p -value).

$$\{y_1 \dots y_m\} \longrightarrow U_m(y_i) \stackrel{?}{=} \frac{i}{m}$$

$$\Rightarrow y < y_i \Leftrightarrow x < x_i$$

$$\Rightarrow \begin{cases} F_n = U_m \\ F(x) = U(y) \end{cases} \quad (\text{se } H_0 \text{ è vera})$$

Il test di k-S può essere anche usato per testare l'uniformità di 2 campionamenti:-

$$x_i \dots x_m \quad F_m(x_i)$$

$$y_i \dots y_M \quad G_m(y_i)$$

$$d_n \equiv \sup |G_m(x) - F_m(x)|$$

$$\text{statistica} \equiv \sqrt{\frac{m \cdot M}{m+M}} \cdot d_n$$

Abbiamo adesso a disposizione vari metodi per confrontare un campionamento con un modello teorico **semplice** -

C'è un test non parametrico, che è **indipendente** dagli altri test: il **run-test**.

In un mazzo di carte ben mescolate, il numero di run è il numero di volte che a un' estrazione il colore della carta è diverso dal colore della precedente. È una misura del "rimiscelamento" del mazzo.

Ad esempio, 2 laboratori hanno fatto misure della massa di una sostanza. Mescolano le loro misure, conto ad esempio quante volte il lab. cambia quando la massa aumenta.

Il numero di run non deve essere né troppo piccolo né troppo grande.

Consideriamo r palle. Ho $M < r$ celle adiacenti. Voglio calcolare il numero di modi in cui posso mettere le r palle (indistinguibili) nelle celle senza lasciarne nessuna vuota -

$\binom{r-1}{M-1}$ Sulla base di questo risultato possiamo progettare il run-test su due popolazioni

α e β -

$\alpha \alpha \alpha \beta \alpha \beta \beta \dots \beta$ ← ordinamento fatto in base ad una grandezza che si è ordinata (la grandezza di cui voglio vedere l'omogeneità, ad es. la massa nell'esempio precedente).

Vedi → normal probability plot -

Tutti i test che abbiamo visto possono essere reinterpretati come **test di decisione**: H_0 + ipotesi alternativa.

15/11/11

		Stati del mondo	
		H_0	H_1
asserisco	H_0	V	α
	H_1	β	V

α \equiv errore di tipo 1 \equiv Loss

β \equiv errore di tipo 2 \equiv contaminazione

Gli errori si dicono anche **loss** e **contaminazione**.

Divido il campione in:

- regione di accettazione di H_0 : $x \in \overline{W_\alpha}$

- regione di rigetto di H_0 (accettazione di H_1): $x \in W_\alpha$

Questo può non essere facile in uno spazio a più dimensioni:

Definizione di W_α :

$$P(x \in W_\alpha | H_0) \equiv \alpha$$

$$P(x \in \overline{W_\alpha} | H_1) \equiv \beta$$

L'ideale sarebbe avere α e β piccolissimi. Questo non si può fare. Quello che si può fare è:

- tenere fisso α
- scegliere W_α in modo da massimizzare la **potenza del test** $1 - \beta = P(x \in W_\alpha | H_1)$.

$W_\alpha \equiv$ **REGIONE CRITICA**

In questo modo possiamo anche progettare il "miglior test".

Lemma di Neyman - Pearson

$$X_1 \dots X_n \sim f(x, \theta)$$

α stabilito ("size" del test, è un numero)

$$H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1$$

W_α tale che $P(\bar{X} \in W_\alpha | H_0) = \alpha$

Si definisce **verosimiglianza** $P(\bar{X} | \theta_{0,1}) \rightarrow L(\bar{X} | \theta_{0,1})$

che in caso di variabili indipendenti si può scrivere sfruttando $P(x) = \prod_i f_i(x_i, \theta_i)$.

Moi lavoriamo con il rapporto tra le verosimiglianze:

$$t = \frac{L(\bar{X} | \theta_1)}{L(\bar{X} | \theta_0)}$$

Massimizziamo la potenza:

$$1 - \beta = \int_{W_\alpha} L(x | \theta_1) dx = \int \frac{L(x, \theta_1)}{L(x, \theta_0)} L(x, \theta_0) dx$$

$$= E_{W_\alpha} \left(\frac{L(x, \theta_1)}{L(x, \theta_0)} \mid H_0 \right)$$

$$t \geq t_\alpha \quad \text{in } W_\alpha$$

$$t < t_\alpha \quad \text{in } \bar{W}_\alpha$$

Se t misurato è $\geq t_\alpha$ rigetto H_0 ed accetto H_1 .
Lo sperimentatore può agire solo nella definizione di α e di W_α .



Questo si può fare solo nel caso in cui H_0 ed H_1 siano semplici (θ_0, θ_1 conosciuti)

Vedi dispense cap. 18.

Esempio: $X_i \sim N(\mu, \sigma^2)$

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 = \sigma_1^2$$

α

$$L = \prod N(x_i, \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2} \frac{\sum (x_i - \mu)^2}{\sigma^2}\right\}$$

$$s^2 \equiv \sum (x_i - \mu)^2$$

$$f = \frac{L(\sigma_1^2, x)}{L(\sigma_0^2, x)} = \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^{n/2} \exp\left\{-\frac{1}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) s^2\right\} \geq f_\alpha$$

Prendiamo il logaritmo:

$$-\frac{n}{2} (\log \sigma_1^2 - \log \sigma_0^2) - \frac{1}{2} \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) s^2 \geq \log f_\alpha$$

Riformulando: $s^2 \geq s_\alpha^2$

$$P(s^2 \geq s_\alpha^2 | H_0) \equiv \alpha$$

Sappiamo che $\frac{s^2}{\sigma_0^2} \sim \chi_m^2 \Rightarrow$ il test è un semplice test del χ^2 .

$P(\mu \geq \mu_\alpha^n | \chi_m^2) = \alpha$ calcolabile con le tabelle.

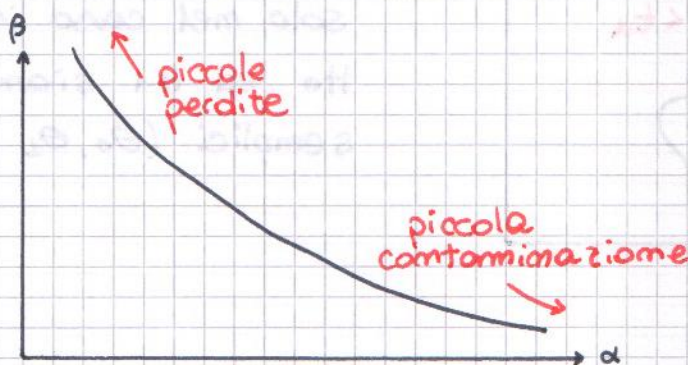
$$\left. \begin{array}{l} \mu \equiv \frac{s^2}{\sigma_0^2} \\ \mu_\alpha^n \equiv s_\alpha^2 \end{array} \right\} \text{ oppure } \left\{ \begin{array}{l} \mu \equiv s^2 \\ \mu_\alpha^n \equiv s_\alpha^2 \sigma_0^2 \end{array} \right.$$

$$1 - \beta = P(s^2 \geq s_\alpha^2 | H_1)$$

Se H_1 è vera, $s^2 \frac{1}{\sigma_1^2} \sim \chi_m^2$.

$$1 - \beta = P\left(\frac{s^2}{\sigma_1^2} \geq \frac{\sigma_0^2 \mu_\alpha^n}{\sigma_1^2}\right)$$

Se $\sigma_1 = \sigma_2$, la potenza del test è α (test a una sola ipotesi).



Curva di ROC (Residual operating characteristic)

Lo scienziato che studia questo tipo di diagrammi è tutt'oggi in pieno sviluppo.

Questo tipo di test non utilizza il singolo esperimento, ma solo una lunga serie di test, tenuti sotto controllo. Il test di significato di Fisher invece consente lo studio del singolo esperimento \Rightarrow è preferito dai fisici.

16/11/11

N.B.: Nel caso ci siano + di 2 ipotesi concorrenti non si può applicare ripetutamente il test di decisione. Occorre preclassificare le ipotesi e scegliere quella che più si adatta ai dati.

I test a una sola ipotesi non possono dare informazioni sulla potenza del test stesso. Non ci sono strumenti per scegliere tra un test e un altro.

\Rightarrow il test decisionale di Neyman - Pearson è matematicamente superiore. Ma non è sempre applicabile in fisica.

H_0 ed H_1 sono simmetriche, ma non completamente (l'ipotesi nulla H_0 è utilizzata per iniziare il test e scegliere α / w_x).

Diagrammi di roc \Rightarrow simmetrizzazione di H_0 e H_1 (sacrificando / ottimizzando perdite e contaminazioni).

Il punto chiave è la verosimiglianza, che è funzione del parametro e dei risultati sperimentali.

$$L(\theta, \vec{x})$$

Se ho $x_1 \dots x_n$ iid $\sim f(x_i | \theta)$, la probabilità congiunta è data da: $P(\vec{x} | \theta) = \prod f(x_i | \theta)$

Si passa da probabilità congiunta a verosimiglianza quando alle variabili aleatorie sostituiamo i risultati delle misure.

Una buona stima del parametro è data da:

$$\hat{\theta} = \sup_{\theta} L(\vec{x}, \theta)$$

$\hat{\theta}$ è una statistica, cioè funzione solo delle misure.

$$\hat{\theta} = \hat{\theta}(\vec{x})$$

Il test decisionale si fa valutando:

$$\Lambda = \frac{L(H_1)}{L(H_0)}$$

Se $\Lambda > \Lambda_\alpha$ (Λ_α da determinare) \Rightarrow rigetto H_0 e scelgo H_1 .

Λ_α si definisce con: $P(\Lambda > \Lambda_\alpha | H_0) = \alpha$ ($\alpha = \text{size del test}$)

Ma pu simmetria c'è anche da considerare:

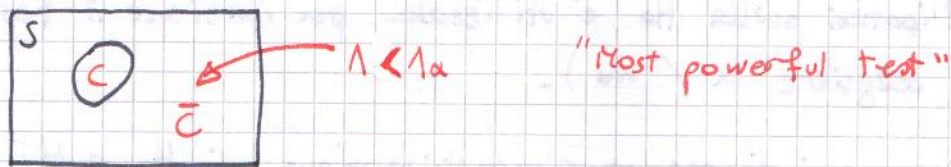
$$P(\text{accettone } H_1 | H_0 \text{ è vera}) = \alpha$$

$$P(\text{accettone } H_0 | H_1 \text{ è vera}) = \beta$$

Come abbiamo impostato il test, β è lasciato libero, e si cerca di minimizzarlo.

$$1 - \beta = P(\text{acc. } H_1 | H_0)$$

Si cerca di massimizzare $1 - \beta$. L'algoritmo di Neyman-Pearson consiste nello scegliere una regione critica C tale che:



Nei casi reali il più delle volte il test non è semplice:

i parametri θ non sono definiti \Rightarrow \nexists most powerful test.

Si continua ad utilizzare il rapporto tra le verosimiglianze, che però adesso dipende anche da parametri non definiti.

Esempio: $H_0 \equiv \{\theta_i\}_0^r = \{\theta_i^0\}_0^r$ con $\theta_i \in \Omega$, $\dim \Omega > r$
 $H_1 \equiv \{\theta_i\}_0^r$ indeterminati (liberi di adattarsi al meglio ai dati)

Ridefinisco Λ :

$$\Lambda \equiv \frac{\sup_{\theta_r, \theta_s} L_0(\theta_r, \theta_s)}{\sup_{\theta_r, \theta_s} L_1(\theta_r, \theta_s)} \quad 0 \leq \Lambda \leq 1$$

Adesso Λ è funzione solo dei dati misurati - è ragionevole usarla come test statistico:

$$\Lambda > \Lambda_c \Rightarrow \text{accetto } H_0$$

Pu scegliere λ_c devo conoscere la distribuzione di λ , oppure usare un montecarlo.

Teorema di Wilks

$$\lambda \equiv -2 \ln \Lambda \sim \chi^2_r$$

(r numero di parametri fissati da H_0)

(supponendo vera H_0)

In questo caso è possibile ridursi ad un test del χ^2 .

17/11/2011

Stima dei parametri

Esempio: $X \sim N(\mu, \sigma^2 = 1)$

Il problema consiste nel trovare la miglior stima di μ che si adatti ai dati stessi.

Uno **stimatore** si indica in genere con $\hat{\mu}$ ed è funzione solo delle misure effettuate (è una **statistica**). Per m (numero delle misure) $\rightarrow \infty$, allora $\hat{\mu}$ dovrebbe tendere a μ . Se questo accade, lo stimatore si dice **consistente**. C'è a volte una piccola distorsione dovuta al campionamento finito. In questo caso:

$$E(\hat{\mu}) = \mu_0 + b(m, \theta_0)$$

b si dice **bias**. Se $b = 0 \Rightarrow \hat{\mu}$ consistente.

Per un singolo valore misurato: $\hat{\mu} \equiv x_i$ (varianza 1)

Per m valori: $\hat{\mu} \equiv \frac{\sum x_i}{m}$ (varianza $\frac{1}{m}$)

$\hat{\mu}$ è una variabile aleatoria.

I metodi più utilizzati per la stima dei parametri sono:

- massima verosimiglianza
- minimi quadrati

Metodo della massima verosimiglianza

Entrambi questi metodi sono stati inventati "ad hoc", non sono ricorati dagli assiomi di Kolmogorov.

\Rightarrow non si applicano in tutti i casi.

Esempio

Urna A: B N N N $P_B = 1/4$

Urna B: B B B N $P_B = 3/4$

Se l'estrazione è fatta con rimpiazzo:

(supponiamo $n = 3$ estrazioni.)

$$P(r \text{ bianche}) = \binom{3}{r} p_B^r (1-p_B)^{3-r}$$

Supponiamo di aver estratto solo palle nere in 3 estrazioni:

$$r = 0$$

$$P(\text{urna A} | r=0) = 27/64$$

$$P(\text{urna B} | r=0) = 4/64$$

Che l'urna sia la A è 27 volte più probabile della B.

Principio di massima verosimiglianza

Dobbiamo scegliere l'ipotesi che massimizza
la probabilità dei dati osservati.

Questo ha qualcosa di Bayesiano.

$$P_{\text{posterior}}(U | \text{Dati}) = \frac{P(\text{Dati} | U) P(U)}{P(\text{Dati})}$$

Non occorre però alcun prior.

N.B.: la verosimiglianza è una probabilità se applicata alle variabili aleatorie, è un numero se applicata ai dati misurati.

$$P_{\text{post}}(U | \text{Dati}) = \frac{P(\text{Dati} | U) P(U)}{\int P(U) P(\text{Dati} | U) dU}$$

Tuttavia, anche in questo caso stiamo facendo un' inferenza induttiva - In matematica \exists solo inferenza deduttiva: tutto quello che viene derivato è già scritto negli assiomi di partenza.

In fisica si fa di più, ma si rischia molto di più di sbagliare.

Vediamo un altro esempio binomiale:

$$P = \binom{m}{r} p^r (1-p)^{m-r}$$

Se in m estrazioni ottengo r successi, qual è p ?

$$L = k p^r (1-p)^{m-r}$$

$$\log L = \log k + r \log p + (m-r) \log(1-p)$$

Massimizzare L = massimizzare $\log L$.

$$\frac{\partial \log L}{\partial p} \equiv 0 \Rightarrow \frac{r}{p} - \frac{m-r}{1-p} = 0 \Rightarrow \hat{p} = \frac{r}{m}$$

Mel caso avessi fatto 2 misure (due serie di estrazioni dalla stessa urna) il metodo funziona allo stesso modo:

x_1 estrazioni $\rightarrow m_1$ successi.

x_2 estrazioni $\rightarrow m_2$ successi.

$$L = k_1 k_2 p^{m_1} (1-p)^{x_1 - m_1} p^{m_2} (1-p)^{x_2 - m_2}$$

Me risulta, banalmente:

$$p_{ML} = \frac{m_1 + m_2}{x_1 + x_2}$$

Esempio

Probabilità Poissoniana. $P(r, \mu) = \frac{e^{-\mu} \mu^r}{r!}$

Supponiamo di aver misurato $r = m$ e di voler stimare μ :

$$L = \frac{e^{-\mu} \mu^m}{m!} \quad \log L = -\mu + m \log \mu - \log m!$$

$$\frac{\partial \log L}{\partial \mu} = -1 + \frac{m}{\mu} \equiv 0 \Rightarrow \hat{\mu} = m$$

(come avevamo già visto)

Esempio m misure iid, $X \sim N(\mu, \sigma^2 \equiv 1)$

$$L = \prod N(\mu, 1) = \left(\frac{1}{2\pi}\right)^{\frac{m}{2}} e^{-\frac{1}{2} \sum (x_i - \mu)^2}$$

$$\log L = -\frac{1}{2} \sum (x_i - \mu)^2$$

$$\frac{\partial \log L}{\partial \mu} = \sum (x_i - \mu) \equiv 0 \Rightarrow m\bar{x} - m\mu = 0$$

$$\hat{\mu} \equiv \bar{x}$$

Se m misure sono state raccolte tutte in una

unica variabile statistica \bar{x} : si dice che

$\hat{\mu} = \bar{x}$ è uno stimatore SUFFICIENTE.

Esempio

Supponiamo di aver fatto n campionamenti

$x_1 \dots x_n$ di una variabile fatta così:

$$f(x) = \frac{3}{8} (1 + \alpha x + x^2)$$

($x \equiv \cos \theta \Rightarrow$ sezione d'urto scattering e^-e^-)

$$L = \prod_i (1 + \alpha x_i + x_i^2)$$

$$\log L = \sum_i \log (1 + \alpha x_i + x_i^2)$$

$$\frac{\partial \log L}{\partial \alpha} = \sum_i \frac{x_i}{1 + \alpha x_i + x_i^2} \equiv 0$$

n soluzioni, non tutte reali, non tutte accettabili.

Per $n \rightarrow \infty$ questa distribuzione tende ad una normale, ma per n finito ci sono molti massimi e molti minimi.

Occorre risolvere graficamente l'equazione e scegliere la soluzione nell'intervallo fisico di interesse.

Esempio

$$x_i \sim N(\mu, \sigma^2) \quad x_1 \dots x_n$$

Vogliamo stimare allo stesso tempo μ e σ^2 .

Risulta:

$$\hat{\mu} = \bar{x}$$
$$\hat{\sigma}^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

Ma $\hat{\sigma}^2$ è distorto, anche se consistente: a denominatore dovrebbe esserci $(n-1)$ e a numeratore μ al posto di $\hat{\mu}$.

In uno spazio dei campioni ad n dimensioni si ha:

$$\hat{\mu} = \bar{x} \quad \hat{\sigma}^2 = \frac{\sum (x_i - \mu)^2}{n}$$

Quando sostituiamo $\hat{\mu} = \bar{x}$ a μ , riduciamo il numero di gradi di libertà da n ad $n-1$.

Esercizio

$$t_1 \dots t_m \quad t \sim \frac{1}{\tau} e^{-t/\tau}$$

$$L(\tau | \vec{t}) = \frac{1}{\tau^m} e^{-\bar{t}m/\tau}$$

$$\log L = -m \log \tau - \frac{\bar{t}m}{\tau}$$

$$\frac{\partial \log L}{\partial \tau} = -\frac{m}{\tau} + \frac{\bar{t}m}{\tau^2} = 0 \Rightarrow \tau = \bar{t}$$

Ma se scriviamo la distribuzione come:

$$t \sim \lambda e^{-\lambda t}$$

$$L(\lambda | \vec{t}) = \lambda^m e^{-\bar{t}m\lambda}$$

$$\frac{\partial \log L}{\partial \lambda} = \frac{m}{\lambda} - \bar{t}m = 0 \Rightarrow \hat{\lambda} = \frac{1}{\bar{t}}$$

Sembra troppo semplice: c'è sicuramente un problema da qualche parte.

Supponiamo $x \sim f(x, \tau)$

Calcolo L dai dati, stimo $\hat{\tau}_{ML}$.

Proprietà di invarianza

$$x \sim f(x, g(\tau)) \Rightarrow \hat{g} = g(\hat{\tau})$$

Esercizio

• calcolare $g(\hat{\tau}) = g(\bar{t})$: $\bar{t}(\hat{\tau}) = \tau$

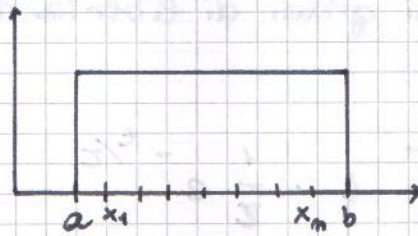
• dato $g(\hat{\tau}) \rightarrow g'(\hat{\lambda})$ risulta $\bar{t}(\hat{\lambda}) = \frac{m}{m-1} \lambda$

Lo stimatore $\hat{\lambda}$ è consistente: $\hat{\lambda} \xrightarrow{m \rightarrow \infty} \lambda$

Ma per m finiti è distorto, va corretto moltiplicandolo per $\frac{m-1}{m}$

Esercizio Prendiamo una distribuzione uniforme.

$$f(x) = \frac{1}{b-a}$$

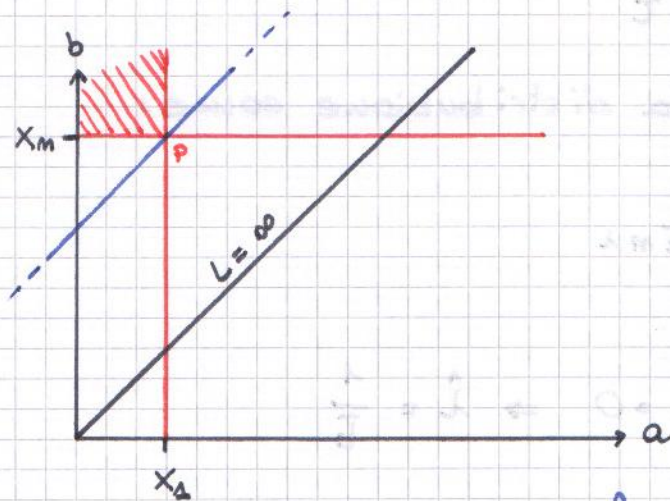


Voglio stimare a e b :

$$L = \left(\frac{1}{b-a} \right)^n$$

Questo non funziona: produrrebbe $a=b$.

Prendiamo la porzione di piano definita da a e b ed ordiniamo i campionamenti per valore crescente.



In questo piano posso definire le curve di uguale verosimiglianza, e sono tutte le rette parallele alla bisettrice (bisettrice $\equiv L = a$).

Deve valere:
$$\begin{cases} a \leq x_{(1)} \\ b \geq x_{(n)} \end{cases}$$

La verosimiglianza massima ottenibile con questi vincoli è la retta che passa per il vertice P.

Abbiamo due stimatori sufficienti:

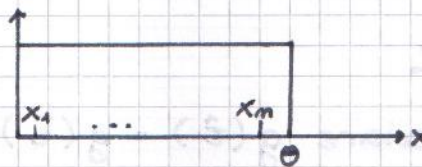
$$\hat{a} = \min_i \{x_i\}$$

$$\hat{b} = \max_i \{x_i\}$$

Esempio

$$f(x) = \frac{1}{\theta} \quad 0 \leq x \leq \theta$$

$$0 \quad x > \theta \vee x < 0$$



Posso scrivere, usando la funzione di Heaviside:

$$f(x, \theta) = \frac{1}{\theta} H(\theta - x)$$

$$L = \prod \left[\frac{1}{\theta} H(\theta - x_i) \right] = \frac{1}{\theta^n} H(\theta - x_{(n)}) \Rightarrow \hat{\theta} = x_{(n)}$$

L'unica misura che definisce il valore stimato di θ è la più grande delle x_i . \Rightarrow Per n finito questo stimatore è distorto.

∃ altri stimatori $\hat{\theta}$?

Il valor medio di qualsiasi x_i è $\theta/2$.

$$E(x_i) = \theta/2 \quad \forall i$$

$$\hat{\theta}'' = 2\bar{x}$$

Paragoniamo i due stimatori: devo scegliere quello a varianza più piccola (il più preciso).

$$\hat{\theta}_1 = x_{[m]} \quad \hat{\theta}_2 = 2\bar{x}$$

Lo stimatore a varianza più piccola sarà lo stimatore più **efficiente**.

Se è presente del rumore di fondo, posso avere conteggi ovunque al di fuori di $[0, \theta]$ (**outlayers**) $\Rightarrow \hat{\theta}_1$ è estremamente fragile. Anche la media tiene conto degli outlayers, ma hanno un peso minore.

$$E(\theta'') = \theta \quad (\hat{\theta}_2 = \theta'')$$

$$\begin{aligned} \text{Var}(\theta'') &= 2 \cdot \text{Var}(\bar{x}) = \\ &= 2 \cdot \text{Var}\left(\frac{\sum x_i}{n}\right) \\ &= \frac{2}{n^2} \sum \text{Var}(x_i) = \frac{2}{n^2} n \frac{\theta^2}{12} = \frac{\theta^2}{6n} \end{aligned}$$

$$\theta'' \sim N\left(\theta, \frac{\theta^2}{6n}\right)$$

$$\hat{\theta}_1 = \theta'$$

$$\begin{aligned} g(\theta') d\theta' &= P[(\theta' \in (\hat{\theta}, \hat{\theta} + d\hat{\theta})) \prod_{i=1}^n (x_i < \hat{\theta})] \\ &= \frac{d\hat{\theta}}{\theta} \left(\frac{\hat{\theta}}{\theta}\right)^{n-1} n = n d\hat{\theta} \frac{\hat{\theta}^{n-1}}{\theta^n} \end{aligned}$$

$$\left(P(x_i < \hat{\theta}) = \frac{\hat{\theta}}{\theta}\right)$$

$$\text{Verificare: } \int_0^\theta g(\hat{\theta}) d\hat{\theta} = 1$$

Possiamo ora calcolare $E(\hat{\theta})$ e $\text{Var}(\hat{\theta})$ - Risultato:

$$E(\hat{\theta}) = \frac{m}{m+1} \theta$$

22/11/2011

Spesso le distribuzioni sono funzioni non solo della variabile aleatoria, ma anche di un set di parametri θ_i .

$$X \sim f(x, \vec{\theta})$$

I θ_i possono essere stimati dalle misure.

Per una variabile NORMALE: $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

In generale, mi aspetto che i punti sperimentali si addensino dove $P(x_1 \dots x_n | \vec{\theta})$ è massimo.

Rovesciando il ragionamento, $L(\vec{\theta} | x_1 \dots x_n)$ è massimo dove $\vec{\theta}$ è il set giusto di parametri

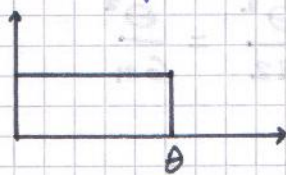
$\hat{\theta}_{LM}$ = stimatore ottenuto col metodo della max verosim.

A volte $\hat{\theta}_{LM}$ non risulta molto in accordo con il "valore vero" $\vec{\theta}$.

Si ha in molti casi: $E(\hat{\theta}_{LM}) = \vec{\theta} + b(\theta)$

Se $b(\theta) \neq 0$, lo stimatore è distorto.

Ad esempio:



$$\hat{\theta}_{ML} = \max \{x_i\}$$
$$E(\hat{\theta}_{ML}) = \theta \frac{m}{m+1}$$

In generale:

$$E(\hat{\theta}_{LM}) = \vec{\theta} + b(\vec{\theta}, m)$$

- $m \rightarrow \infty \quad b \rightarrow 0 \Rightarrow \hat{\theta}$ è CONSISTENTE
- $b \neq 0 \Rightarrow \hat{\theta}$ è NON DISTORTO

Esempio: $t_1 \dots t_m \sim \frac{1}{\tau} e^{-t/\tau} \quad \hat{\tau} = \bar{t}$

se passiamo da τ ad $1/\lambda$: $\hat{\lambda} = \frac{1}{\hat{\tau}} = \frac{1}{\bar{t}}$

(PROPRIETÀ DI INVARIANZA)

Torniamo all'esempio della distribuzione uniforme.

$$\hat{\theta}_{ML} \equiv \max \{x_i\} \frac{n+1}{n}$$

$$\hat{\theta}_0 \equiv 2\bar{x}$$

Usando il TLC ricaviamo che:

$$\text{Var}(x_i) = \frac{\theta^2}{12}$$

$$\text{Var}(\bar{x}) = \frac{12\theta^2}{n}$$

$$\Rightarrow \hat{\theta}_0 \sim N\left(\theta, \frac{12\theta^2}{n}\right)$$

Invece:

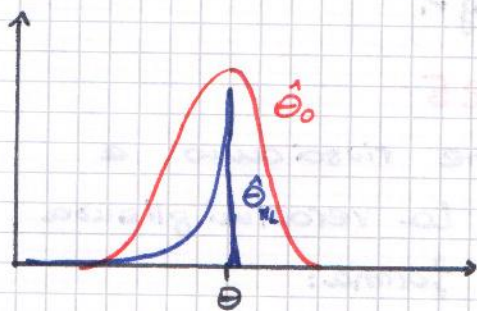
$$\hat{\theta}_{ML} \sim n \frac{\hat{\theta}^{n-1}}{\theta^n} \frac{n+1}{n} = (n+1) \frac{\hat{\theta}^{n-1}}{\theta^n}$$

$$\text{Var}(\hat{\theta}_{ML}) = \theta^2 \frac{n}{(n+1)^2(n+2)}$$

→ va a θ più velocemente di $1/n \Rightarrow$ questo è

uno stimatore

SUPERCONVERGENTE.



Criterio per scegliere uno stimatore

✗ un criterio assoluto. Lo stimatore deve essere:

- non distorto
- con la varianza più piccola possibile.

Da un punto di vista pratico si preferisce invece **minimizzare l'errore quadratico minimo.**

$$\begin{aligned} \text{mse} &= E((\theta - T)^2) = E((\theta - E(T) + E(T) - T)^2) \\ &= E((\theta - E(T))^2) + \underbrace{E((E(T) - T)^2)}_{= \text{Var}(T)} + E[(\theta - E(T))(T - E(T))] \end{aligned}$$

(con T stimatore di θ).

Minimizzando mse riusciamo a minimizzare la varianza dello stimatore, a costo di un piccolo bias.

Supponiamo di avere $x \sim f(x, \theta)$ n iid

$$L = \prod_{i=1}^n f(x_i, \theta)$$

Dipende dall'esperimento stabilire se n è fisso.

Supponiamo che sia una variabile aleatoria $n(\theta)$.

Dobbiamo modificare la verosimiglianza:

$$L \equiv P(n) \prod_{i=1}^n f(x_i, \theta) \equiv \frac{e^{-\mu} \mu^n}{n!} \prod_{i=1}^n f(x_i, \theta), \quad \mu = \mu(\theta)$$

Se abbiamo dati binari in un istogramma:

$$n_{TOT} \rightarrow m_1 \dots m_L \quad n_{TOT} = \sum_i m_i$$

$$P(m_1 \dots m_L) = \frac{n_{TOT}!}{m_1! \dots m_L!} p_1^{m_1} \dots p_L^{m_L}$$

$$P_i = \int_C f(x_i, \theta) dx$$

\Rightarrow va massimizzato $\log L = \sum_i m_i \log P_i$

SUFFICIENZA DI UNO STIMATORE

x_i iid $f(x_i, \theta)$

Può darsi che riusciamo a
fattorizzare la verosimiglianza
in questa forma:

$$L = \prod_{i=1}^n f(x_i, \theta)$$

$$L = g(\hat{\theta}, \theta) P_i(x) \quad \text{Teorema di fattorizzazione di Neyman - Fisher}$$

g dipende ancora dalle x_i
ma solo attraverso $\hat{\theta}$

Se questo accade, si dice che $\hat{\theta}$ è uno stimatore
sufficiente per θ .

Prima di questo teorema, la definizione di sufficienza
era un'altra:

$$P(\bar{X} = \bar{x} \mid \hat{\theta}(\bar{x}) = t) = m(x)$$

Queste due definizioni sono equivalenti.

23/11/2011

$L = g(\hat{\theta}, \theta) h(x)$ *Condizione di fattorizzazione di Neyman-Pearson*

Traue casi particolari (se il supporto di x non dipende da θ), g è la densità di probabilità di $\hat{\theta}$.

Quando questa condizione è verificata, l'equazione

$$\frac{\partial \log L}{\partial \hat{\theta}} = 0$$

si riduce a

$$\frac{\partial g}{\partial \hat{\theta}} = 0$$

Supponiamo di essere nella condizione $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$:

$$\frac{\partial^2 \log L}{\partial \theta \partial x_i} = \frac{\partial^2 \log g}{\partial \theta \partial x_i} = \frac{\partial^2 \log g}{\partial \theta \partial \hat{\theta}} \frac{\partial \hat{\theta}}{\partial x_i}$$

$$\frac{\frac{\partial^2 \log g}{\partial \theta \partial x_i}}{\frac{\partial^2 \log g}{\partial \theta \partial x_j}} = k(x) \Rightarrow \frac{\partial^2 \log g}{\partial \theta \partial x} = Q_1(\theta) Q_2(x)$$

Integrando troviamo:

$$\log g = g_1(\theta) \cdot g_2(x) + k_1(x) + h(\theta)$$

*Uno stimatore sufficiente è uno stimatore efficiente -
⇒ il metodo della max verosimiglianza lo trova.*

Uno stimatore efficiente è lo stimatore a varianza più piccola ottenibile, compatibilmente con la distribuzione in esame.

Ad es.: gaussiana

media campionaria = efficiente

mediana = eff. 50%, ovvero varianza doppia
(ma più robusto)

Disuguaglianza di Schwartz

$$E((tX+Y)^2) \geq 0$$

$$E(t^2x^2 + y^2 + 2txy) \geq 0$$

$$\Rightarrow E(xy) \leq E(x^2)E(y^2)$$

$$y = kx \Rightarrow E(xy) = E(x^2)E(y^2)$$

Supponiamo $t(\vec{x})$ stimatore di θ .

$$E(t) = \int_{\Omega} L(\vec{x}, \theta) t d\vec{x} = \theta + b(\theta)$$

Consideriamo il caso $\Omega \neq \Omega(\theta)$.

(Ω = supporto delle x_i)

$$\int_{\Omega} L(\vec{x}, \theta) d\vec{x} = 1 \quad \text{Condizione di normalizzazione}$$

$$\int t \frac{dL}{d\theta} dV = 1 + \frac{db}{d\theta}$$

È conveniente usare la derivata logaritmica:

$$\frac{\partial L}{\partial \theta} = \frac{\partial \log L}{\partial \theta} L$$

$$\int t \frac{d \log L}{d\theta} L dV = 1 + \frac{db}{d\theta}$$

$$\int \frac{d \log L}{d\theta} dVL = 0$$

$$\int (t - \theta - b) \frac{\partial \log L}{\partial \theta} L dV = 1 + \frac{db}{d\theta}$$

$$\int (t - \theta - b) \sqrt{L} \cdot \sqrt{L} \frac{\partial \log L}{\partial \theta} dV = 1 + \frac{\partial b}{\partial \theta}$$

$$\int (t - \theta - b)^2 L dV = \text{Var}(t) \cdot \int \left(\frac{\partial \log L}{\partial \theta} \right)^2 L dV = \left(1 + \frac{\partial b}{\partial \theta} \right)^2$$

Utilizziamo la disuguaglianza:

$$\text{Var}(t) \geq \frac{\left(1 + \frac{db}{d\theta} \right)^2}{E\left(\left(\frac{\partial \log L}{\partial \theta} \right)^2 \right)}$$

EFFICIENZA \Leftrightarrow vale l' "=" -

La condizione di efficienza è data da:

$$\frac{\partial \log L}{\partial \theta} = A(\theta)(t - \theta - b) \Leftrightarrow t \text{ efficiente}$$

Errore nelle dispense sull' MSR:

pag. 216 il - è un + (TLC)

Deriviamo la normalizzazione rispetto a θ :

$$\int dV L = 1$$

$$\int \frac{d \log L}{d \theta} L dV = 0 \Rightarrow E\left(\frac{d \log L}{d \theta}\right) = 0$$

$$E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) = -E\left(\left(\frac{\partial \log L}{\partial \theta}\right)^2\right)$$

$$\text{Var}(t) = \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{E\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right)}$$

Nel caso che $b = 0$, $A(\theta)$ è l'inverso della varianza.

riscriviamo meglio questa roba:

$$\int L dV = 1 \quad \text{lo derivo due volte e trovo}$$

$$\int \frac{\partial L}{\partial \theta} = 0 \quad \rightarrow \text{sost. il log} \rightarrow E\left(\frac{\partial \log L}{\partial \theta}\right) = 0$$

derivo quello sopra e trovo

$$\frac{d}{d\theta} \left[\int \frac{\partial \log L}{\partial \theta} L dV \right] = 0 \Rightarrow E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right) = -E\left(\frac{\partial \log L}{\partial \theta}\right)^2$$

$$\Rightarrow \text{Var}(\hat{\theta}) \approx \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{E\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right)}$$

24/11/2011

questo mi permette di calibrare gli esperimenti \rightarrow voglio una certa Var sul parametro.

per $b = 0$ posso scrivere $L = \prod F(x_i)$ \rightarrow facendo la der. 2^a esce θ e il val. di aspett., siccome gli $E\left(\frac{\partial^2 F}{\partial \theta^2}\right)$ sono uguali mi esce fuori un n

= n misure. se voglio una certa $\text{Var}(\hat{\theta})$ posso avere un'idea di quante mis fare (E può dip. da θ , che non so, ma che \rightarrow ~~sta~~ ci butterò dentro un valore teorico o cose così).

Caso di est. e FF. $\frac{\partial \log L}{\partial \theta} = A(\theta)(t - \theta - b)$
 vediamo cos'è A $(t = \hat{\theta})$

$$\Rightarrow E\left((\hat{\theta} - \theta - b) \frac{\partial \log L}{\partial \theta}\right) = A E\left((\hat{\theta} - \theta - b)^2\right)$$

supp. non ci sia distorsione: $\frac{db}{d\theta} = 0$

$$\Rightarrow A(\theta) = \frac{1}{\text{Var}(\hat{\theta})}$$

Derivo l'espressione per $\frac{\partial \log L}{\partial \theta}$

$$\frac{\partial^2 \log L}{\partial \theta^2} \Big|_{\hat{\theta}} = \frac{\partial A(\theta)}{\partial \theta} (\hat{\theta} - \theta) - A(\theta) \approx$$

\uparrow $\hat{\theta}$
 sup. $= 0$

↳ questa $\bar{e} < 0 \Rightarrow$ allora tutte le sol. con $\hat{\theta}_i$ sono max.

Ma una sol. reg. tra due max deve avere un min \rightarrow non c'è, quindi la sol. \bar{e} unica.

es. distrib Norm. n val x_i :

$$L = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{1}{2} \sum (x_i - \mu)^2}$$

$$\frac{\partial \log L}{\partial \mu} = - \frac{\partial}{\partial \mu} \left(\frac{1}{2} \sum (x_i - \mu)^2 \right) = n\bar{x} - n\mu =$$

$$= n(\bar{x} - \mu)$$

ho $\hat{\mu}$ trovato la cond. di efficienza!

$$\Rightarrow \bar{x} \bar{e} \text{ eff.}$$

• r poiss.

$$L = \prod \frac{1}{r!} e^{-\mu} \mu^r \Rightarrow \log L$$

$$\Rightarrow \frac{\partial \log L}{\partial \mu} = -n + \frac{n\bar{r}}{\mu} = \frac{n}{\mu} (\bar{r} - \mu)$$

$$\Rightarrow \bar{r} \bar{e} \text{ eff.}$$

Consideriamo quella che chiamiamo il det. della
distrib. di prima: $I = E \left(- \frac{\partial^2 \log L}{\partial \theta^2} \right)$ (si chiama ^{informaz.} ~~var.~~
di Fisher)

se $X \sim N(\mu, \sigma^2)$

$$L = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \quad I_x = E \left(\frac{(x-\mu)^2}{\sigma^2} \right) = \frac{1}{\sigma^2}$$

\Rightarrow
n misure: $I_n = n I_x$

• Se ho 2 parametri, posso scrivere I così:

$$I_x(\theta_1, \theta_2) = E \left(\frac{\partial \log L}{\partial \theta_1}, \frac{\partial \log L}{\partial \theta_2} \right) = - E \left(\frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_2} \right)$$

\rightarrow per $n \rightarrow \infty$ la distrib. \rightarrow Norm. quindi gli
estim. \rightarrow estim. efficienti

\Rightarrow posso calc. a priori (con prob.?)

es. ∇ Calcolare I_x (e $\sim I_x(\mu, \sigma^2)$) $X \sim N(\mu, \sigma^2)$

es. $\mu^+ \rightarrow e^+ + v_m + v_c$

$\frac{d\sigma}{d\alpha} = 1 + \alpha \cos \sigma$ chiamo $\cos \sigma = x \rightarrow$ distrib.

\Rightarrow impongo norm. (integrale sull'accettanza = 1)

$$f(x, \alpha) = K(1 + \alpha x)$$

$$(\text{Var}(\alpha))^{-1} = E \left(\frac{\partial^2 \log L}{\partial \alpha^2} \right)$$

$$L = 1 + \alpha x = F$$

$$E \left(\frac{\partial^2 \log L}{\partial \alpha^2} \right) = E \left(- \frac{\partial}{\partial \alpha} \left(\frac{1}{F} \frac{\partial F}{\partial \alpha} \right) \right) = E \left(- \frac{1}{F^2} \left(\frac{\partial F}{\partial \alpha} \right)^2 + \frac{1}{F} \frac{\partial^2 F}{\partial \alpha^2} \right) =$$

$$= \int \frac{1}{F} \left(\frac{\partial F}{\partial \alpha} \right)^2 dx$$

poi integro $\frac{\partial}{\partial \alpha}$ va fuori molli per F

-> se ho n misure $\text{Var}(\bar{x})^{-1} = n \text{E}(\dots)$
 per scegliere dei val. di α so quanti n
 devo prendere per avere $\text{Var} = \text{tot}$.

es. $t \sim \frac{1}{2} e^{-\frac{t}{\tau}}$ distrib. ~~tempo di deca~~
 tempi di arrivo raggi cosmici (?)

ma io non misuro all'inf. -> misuro per
 un certo T poi ho un tot. di eventi che
 cascano fuori -> overflow
 allora $t \in [0, T]$

=> la distrib. giusta è

$$\frac{1}{2} e^{-t/\tau} \frac{1}{1 - e^{-T/\tau}} \quad \text{tra } 0 \text{ e } T$$

↑
questa è norm.

Se però uso solo questa cosa, non
 ho messo nella verosim. il fatto che
 in un certo num. di eventi sono andati
 in overflow => non ho estim. eff.

$$P(\text{ovf}) = P(t > T | \tau) = e^{-T/\tau}$$

$$\rightarrow P(n_{\text{ovf}}, p_{\text{ovf}}, n) = \binom{n}{n_{\text{ovf}}} \cdot p_{\text{ovf}}^{n_{\text{ovf}}} (1 - p_{\text{ovf}})^{n - n_{\text{ovf}}}$$

* eventi

↑
 questo devo
 moltiplicarlo
 per L => così tengo
 conto di tutto

$$\text{se } n_{\text{ovf}} = 0 \Rightarrow (1 - p_{\text{ovf}})^n$$

si sempl. con il der. $\frac{1}{1 - e^{-T/\tau}} \Rightarrow$ torno da capo

⇒ se no L completa \bar{e} :

$$L = \binom{n}{n_{\text{ovf}}} e^{-n_{\text{ovf}} T/\tau} \left(\frac{e^{-\bar{e}/\tau}}{\tau} \right)^{n - n_{\text{ovf}}}$$

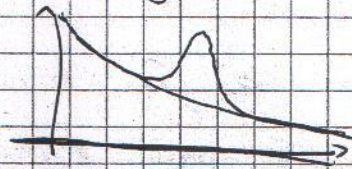
es. prova a simularlo ⇒ root da l'exp. già fatto.

Prendi $n = 10$ $\tau = 1$ $\bar{E} = 2$ (?)

- 2. DISTRIB -

Un Fondo con un picco → 2 famiglie di distrib.

es. una gauss su un exp



$-g(\theta_1)$ (exp.)

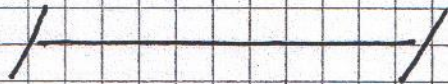
$(0 < d < 1)$

$f_h(\mu, \sigma^2)$

$$L = d g(\theta) + (1-d) f_h(\mu, \sigma^2)$$

massimizzando L sempre ??

$\left\{ \begin{array}{l} \text{Valore di } d, \text{ ovvero quanto contano} \\ \text{una e l'altra, poi posso verificare} \\ \text{se } f_h \text{ \u00e9 comp. con zero o no.} \end{array} \right.$



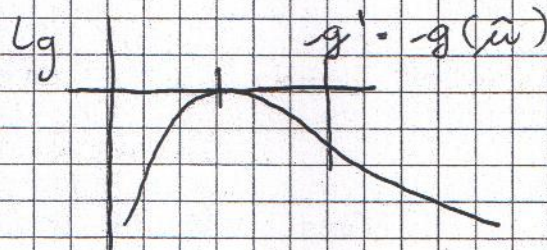
INVARIANZA

$\theta \rightarrow g(\theta)$ bigettiva

$$P(x, \theta_0) = P(x, g(\theta_0))$$

$$L_{\theta}(\theta_0 | x) = L_g(g(\theta_0) | x)$$

sopp. L sia fatta così:



↑
ho trovato $\hat{\mu}$
estim. del max di L ⇒ $\hat{\mu}(g)$

$$L(g') = L(\hat{\mu}) > L(\theta) = L_g(-g)$$

$$\Rightarrow g' = \hat{g} \quad \hat{g} = g(\hat{\mu})$$

nell'ip. però, che gli estimi non siano
biasati.

\Rightarrow Disug. di Jensen (vale per funz. concesse)

$$g(E(x)) \leq E(g(x))$$

allora se $\hat{\mu}$ era non biasato

\hat{g} lo è per forza. (es. es. distrib.

exp. \Rightarrow se $\hat{\tau}$ è non biasato, $\hat{\tau} = \frac{1}{n} \sum_{i=1}^n \tau_i$ lo

$$\bar{\tau} \Rightarrow E(\hat{\tau}) = \tau \quad E(\hat{\tau}) = \frac{n}{n-1} \tau \cdot \frac{n-1}{n} \quad \left(\tau = \frac{1}{2} \right)$$



15/10/11

"Divergenza" \rightarrow principio di Verosimiglianza

Se ho 2 exp. che stimano dati par. θ .

$E_1 \rightarrow \theta_1$ $E_2 \rightarrow \theta_2$ se $L_{E_1} = L_{E_2}$ (o anche solo prop. allora qualsiasi inferenza faccio su θ_1 devo farla anche su θ_2 .

es. $n = 12$ lanci $T = 3$ teste.

ma ho 3 diste. possibili

se

• n fisso $\rightarrow T \sim \text{bin.}$ ~~indist.~~

• se T fisso $\rightarrow n \sim \text{Pascal}$

• $L = p^r (1-p)^{n-r}$ questa è la Bayesiana che si basa solo sulla misura effettivamente fatta, non sulla distrib. associata

• test. di signif.

$$\begin{aligned} \xrightarrow{1^\circ \text{ caso}} P(r=0, 1, 2, 3 \mid H_0: p=1/2, n=12) &= \left(\frac{1}{2}\right)^{12} \left(\binom{12}{0} + \binom{12}{1} + \binom{12}{2} + \binom{12}{3} \right) \\ &= 7,3\% \end{aligned}$$

$$\xrightarrow{2^\circ \text{ caso}} P(n \mid r, p) = \binom{n-1}{r-1} p^r (1-p)^{n-r} = 3,2\%$$

questa cosa è perché le code sono diverse.

\Rightarrow allora, su due casi che danno LA STESSA L ho n risultati diversi.

Perché L si basa sui dati presi.

Signif. si basa ~~si~~ anche su misure che potevo prendere ma non ho preso (???)

METODO DEGLI INTERVALLI DI CONFIDENZA

• sia $X \sim N(\mu, \sigma^2)$

$$X \rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$P(Z \in [z_1, z_2]) = \alpha$ prob. di stare

in un intervallo \Rightarrow può essere qualunque
ma conviene prenderlo simmetrico

\rightarrow si scrive come:

$$P(Z \leq z_1) = \frac{1-\alpha}{2} \quad P(Z \geq z_2) = \frac{1-\alpha}{2}$$

(in tot deve fare 1)

se considero

$$P(Z \in [-1, 1]) = 68,3\% \text{ (per la gauss.)}$$

$$\hookrightarrow P(-1 \leq Z \leq 1) = P(X - \sigma \leq \mu \leq X + \sigma)$$

\uparrow

ho prob. che μ stia in un
intervallo finito.

\Rightarrow Ho info su μ e su dove sta.

Però a questo punto ho fissato

un intervallo finito con certe caratt.

\Rightarrow non parlo di prob. dell'intervallo ma

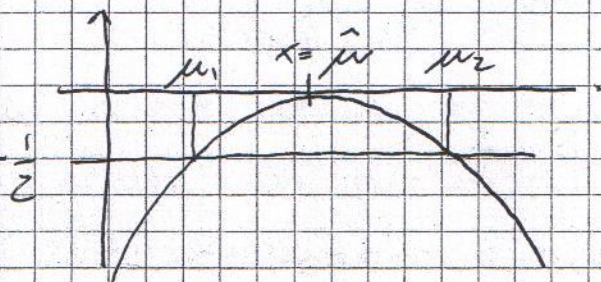
di CONFIDENZA " "

\rightarrow INTERVALLO DI CONFIDENZA

Nota: se ho $P(X - z_2 \sigma \leq \mu \leq X + z_2 \sigma) = \alpha$
dato α trovo z_2 , ovvero l'intervallo
di confidenza

Prendiamo il max della ver.

$$\log L = -\frac{(x-\mu)^2}{2\sigma^2}$$



=> determino graficamente l'int. di conf.

es. (con $-\frac{1}{2}$ è il 68%)

(imp. $\log L = \log \alpha$??)
o cose così ??

Se ho $\log L$ che non è gaussiano, e so che $\exists g(\theta)$ che me lo fa diventare gaussiano (ovvero $\log L$ quadratico) (ovvero $g(\theta) \sim N(\mu, \sigma)$)

Posso usarlo per trovare l'int. di confid.

ovvero: cerco θ trovo θ_1 e θ_2 t.c.

$$g_1 \text{ e } g_2 \text{ t.c. } [L(g_1) = L(g_2) = -\frac{1}{2} (??)]$$

t.c. Facciamo la parte di μ_1 e μ_2

=> allora so che quell'int. ha il 68% di confidenza

Ancora sull'invarianza

L'importante è che ~~se~~ posso scegliere di minimizzare
risp. a una funzione del par. che rende più
facile il conto (es. gaussiana con $g = e^{-x^2}$)

es. $X \sim \text{Bin}(p, N)$

$$X \sim \binom{N}{r} p^r (1-p)^{N-r} \Rightarrow \hat{p} = \frac{r}{N}$$

$$E(\hat{p}) = \frac{E(r)}{N} = \frac{Np}{N} = p \Rightarrow \text{NON DISTORTO}$$

se considero:

$$\hat{p}^2 = \left(\frac{r}{n}\right)^2 \quad E(\hat{p}^2) \rightarrow p^2$$

$$\text{ma } E(\hat{p}^2) = \frac{E(r^2)}{N} = p^2 + \frac{p}{N}$$

è distorto

es. \rightarrow Correggere

$$\hat{p}_{ML}^2 \text{ in modo che } E(\hat{p}^2) = p^2$$

(sarà un polinomio in $\frac{r}{N}$)

Con il metodo ML posso determinare

Facilmente l'INTERVALLO DI CONFIDENZA:

sia $X \sim F(x, \theta)$ voglio un intervallo
($\theta_1(x)$, $\theta_2(x)$) di s. def. sull'asse reale.

Facendo l'esp. voglio ho θ_1 e θ_2 fissi.

Voglio che $I_x = \{\theta_1, \theta_2\}$ sia t.c. $P(\theta \in I_x) = \alpha$

è int. funz. var. al.

con α definito

all'inizio \Rightarrow Va bene parlare di probabilità

perché θ è fisso, ma θ_1 e θ_2 sono funz.

c'è $x \Rightarrow$ si parla di prob. di copertura

\Rightarrow prob. che I_x copra θ .

\Rightarrow (Misurando!) ho degli intervalli \Rightarrow la frequenza con

Cui gli intervalli coprono θ $\bar{\epsilon} = \alpha$ o non inferiore ad α .

$\Rightarrow \theta$ $\bar{\epsilon}$ ignoto, ma posso costruire gli intervalli OTTENUTI DALLE MISURE in modo che ci sia "prob." α di avere θ lì dentro \Rightarrow siccome sono valori misurati non parlo di prob. ma di LIVELLO DI CONFIDENZA.

Si riescono a costruire ^(anche se non so θ) perché le distr. di X hanno traccia del par θ .

es considero $X \sim N(\mu, \sigma^2)$ (supp. di sapere σ^2 e cercare μ)
 $\Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ \rightarrow faccio finta di misurare $Z \Rightarrow$ ho nascosto i parametri.
considero:

$$P(Z \geq z_1) \quad P(Z \leq z_2)$$

$P(z_2 \leq Z \leq z_1) = \alpha$ questi intervalli hanno contenuto di prob. def. da $N(0, 1)$

\downarrow
per fare questi integrali conviene spesso conviene usare $1 - \alpha$ come di altre volte scelgo valori tipo 90% o 99%

usare metodi numerici.

\rightarrow NOTA: in gen. z_1 e z_2 non sono simm.

• spesso si usa z_1 o $z_2 = \pm z$

• altri molto comuni sono invece quelli simm

~~Prendiamo~~ Vale una la seguente:

$$P(Z \leq z_1) = \frac{1 - \alpha}{2} = P(Z \geq z_2)$$

$\Rightarrow \bar{\epsilon}$ simmetrico

Riscriviamo $P(z_2 \leq \frac{x-\mu}{\sigma} \leq z_1)$

come: $P(x - z_1 \sigma \leq \mu \leq x - z_2 \sigma) = d$

$$\Rightarrow \underline{I}_{CL=d} = \{x - z_1 \sigma, x - z_2 \sigma\}$$

Dal conto con la z trovo z_1 e z_2

\Rightarrow per ogni misura x_i mi dà un

int. $I_{x_i} \Rightarrow$ ha proba d su n misure

ha l'int. x_i conterrà μ di valore

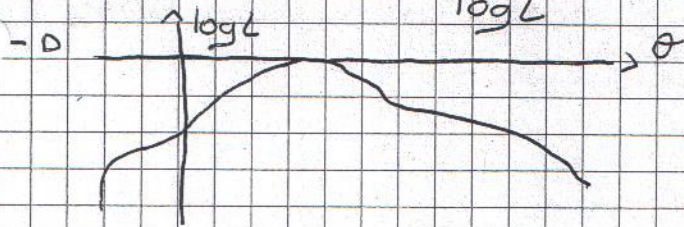
con frequenza d .

\rightarrow Ma non saprò mai se il specifico intervallo becca d o no.

Questo funziona bene quando so la distribuzione, ma se la d sui casi non parametrici funziona meno bene.

Vediamolo con applicata l'invarianza.

Prendiamo un $\log L$ fatto così (va anche con più massimi)



Supp. che $\exists g$ t.c. $g \in \mathcal{L}_g(g, x)$ e sia normale

$$\Rightarrow g(\theta) \cdot \log L_g(g, x) = -\frac{1}{2} \frac{(g - \hat{g})^2}{\sigma_g^2}$$

ha che: $L_g(g(\theta_0), x) = L_g(\theta_0, x)$

$$e \hat{g}(\theta) = g(\hat{\theta})$$

(per un dato θ_0 o

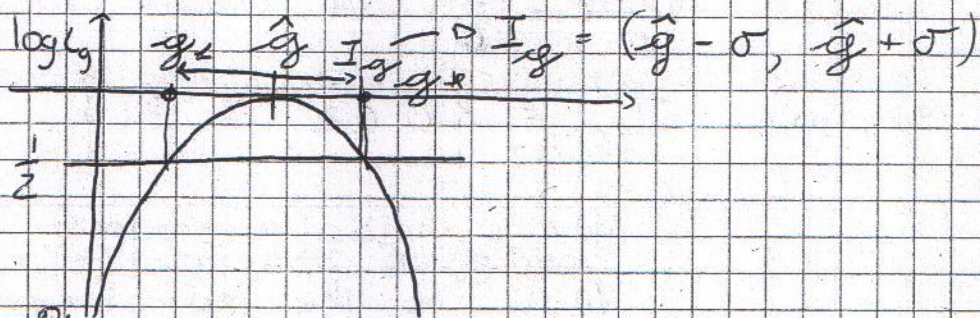
per il max?)

Prendiamo $I_g = (x - \sigma z_1, x + \sigma z_2)$

per comodità prendiamo $\alpha = 68,3 \dots \% \Rightarrow z_1 = z_2 = 1$

I_g si calcola partendo da L_g e mettendo

$\log L_g = \frac{1}{2}$ e intersecando



Ai punti g_L e g_R
corrispondono θ_L e θ_R

$\Rightarrow \theta_L = g^{-1}(g_L) \Rightarrow$ così non ho bisogno di vedere tutta la g^{-1} per l'intero intervallo

Tutte le inferenze valide per I_g si ~~possono~~ sono essere valide per I_θ

$$\Rightarrow \log L(\theta_R) = \log L(\theta_L) = \log L(\hat{\theta}) \cdot \frac{1}{2}$$

\Rightarrow Qualunque sia la forma di $\log L_\theta$, basta che tracci la linea a $\frac{1}{2}$ e trovo l'intervallo di conf. al 68%; idem per altri valori.

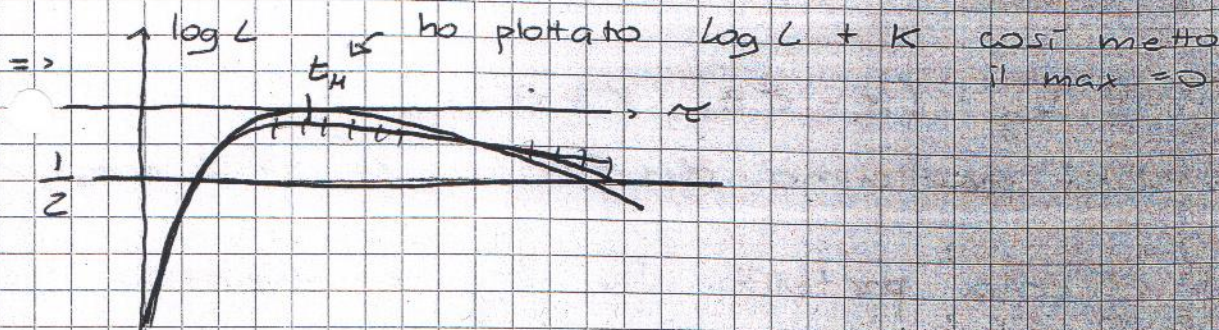
Se $g \neq \hat{\theta}$, allora sarà solo un metodo approssimato.

\rightarrow es \rightarrow distrib. exp.

$t \sim e^{-t/\tau}$ ho fatto una sola misura. t_m

$$L(\tau, t_m) = \frac{1}{\tau} e^{-t_m/\tau}$$

$$\Rightarrow \log L(\tau, t_m) = -\log \tau - \frac{t_m}{\tau}$$



$$\log L = -\log r - \frac{t_M}{2} + 1 \sqrt{K}$$

=> trovo $I_{68\%} = \{0,42; 3,31\}$

quando inizio a trovarli asimmetrici
 vuol dire che ~~non~~ sono stimati
 maluccio.

=> facciamo il confronto con gli int. esatti

calcolando $P(z \leq z_1) = \int_0^{z_1} e^{-z} dz$
 con $z = \frac{t}{\sigma}$

→ come l'altro già fatto

o voglio int. centrale => uguale prob. sulle code,

ovvero: $P(z \leq z_1) = \frac{1-\alpha}{2} = P(z \geq z_2)$

$$\int_0^{z_1} e^{-z} dz = \frac{1-\alpha}{2} \Rightarrow 1 - e^{-z_1} = -\frac{\alpha}{2} + 1$$

$$\Rightarrow z_1 = -\log\left(\frac{1+\alpha}{2}\right) \quad z_2 = -\log\left(\frac{1-\alpha}{2}\right)$$

=> viene $I_{68\% \text{ ESATTO}} = \{0,55; 5,74\}$

=> ~~g~~! era

solo un'approx.

=> con più misure devo usare t che con tante

misure $L \rightarrow N$ form. da se' => la
 situazione ^{diventa} $\frac{1}{\sqrt{2}}$ più favorevole.

-> NON USARE QUESTO METODO PER TEST. D'IPOTESI

es. $\theta = 0,5$ al 68% di confidenza va bene?

=> allora deve essere incluso nell'int. => risente

tanto dell'errore sull'intervallo.

-> però si dice mentre prendo i dati per dare un'idea di cosa sto combinando, non importa se sbaglia anche di un fattore 2.

-> Proprietà estimatori => normalità asintotica

$$(\hat{\theta} - \theta_0) \sqrt{\frac{J_n}{n}} \xrightarrow[n \rightarrow \infty]{\text{distr.}} N(0, 1) \quad \leftarrow \text{PROPR. DI NORM. ASINT.}$$

la dim. di questa cosa è sulle disperse

J_n = estim. dell'inform. di Fisher

$$I_x = - E \left(\frac{\partial^2 \log L}{\partial \theta^2} \right) \approx - \frac{1}{n} E \left. \frac{\partial^2 \log F(x, \theta)}{\partial \theta^2} \right|_{\hat{\theta}}$$

= $\frac{J_n}{n}$

↳ sto usando legge dei grandi numeri imponendo che sia la media campionaria che $\hat{\theta} \rightarrow \theta_0$ tendano a θ_0

PROPRIETÀ EST. COM ML:

- ~~sono~~ consistenti

- sono efficienti \Leftrightarrow sufficienti \Rightarrow le trova in automatico

ma invece gli est. non bias distorti questo metodo non li trova \rightarrow devo fare ulteriori conti.

30/11/2011

Altri metodi per calcolare gli int. di confidenza

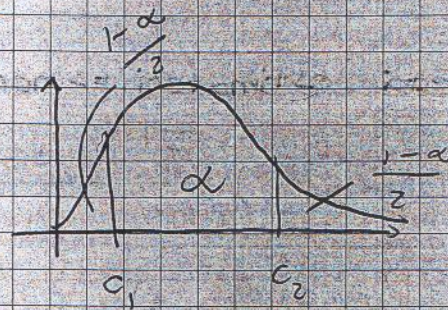
• Var. che possono essere ~~ma~~ si possono trasformare in modo da avere una var. Z senza parametri con cui calcolare gli int. di confid. \rightarrow metodo di prima.

es. se sulla distr. di prima devo stim. la

σ invece della μ .

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$\mu_{\text{est}} = \frac{S^2 (n-1)}{\sigma^2} \sim \chi^2_{n-1}$$



\Rightarrow prendo int. simm.

$$P(c_1 \leq \mu \leq c_2) = \alpha$$

si calcola numericamente

$$\Rightarrow P\left(c_1 \leq \frac{S^2 (n-1)}{\sigma^2} \leq c_2\right) = P\left(\frac{S^2 (n-1)}{c_2} \leq \sigma^2 \leq \frac{S^2 (n-1)}{c_1}\right) = \alpha$$

c) il max di L , a parte questo non ho in automatico info ~~sub~~ ~~part~~ se il modello si accorda o no, devo fare test d'ipotesi.

es. $X_i \sim e^{-t/\tau} \frac{1}{\tau}$ n $\bar{X} = \frac{\sum X_i}{n}$

Verifica che $\frac{2n\bar{X}}{\tau} \sim \chi^2_{2n}$

$$\Rightarrow \text{da qui si trova } P\left(\frac{2n\bar{X}}{c_2} \leq \tau \leq \frac{2n\bar{X}}{c_1}\right) = \alpha$$

Consideriamo 2 variabili $X \sim N(\mu_x, \sigma^2)$ n varianze uguali e
 $Y \sim N(\mu_y, \sigma^2)$ m non note

$$\bar{X} = \frac{\sum x_i}{n} \quad \bar{Y} = \frac{\sum y_i}{m}$$

$$S_x^2 = \frac{\sum (x_i - \bar{X})^2}{n-1}$$

$$S_y^2 = \frac{\sum (y_i - \bar{Y})^2}{m-1}$$

se σ sono uguali uso stimatore $S^2 = \frac{S_x^2(n-1) + S_y^2(m-1)}{n+m-2}$

se ho il dubbio che siano diverse

devo prenderli separati

$$\Rightarrow \text{student } t = \frac{\mu_x - \mu_y + (\bar{x} - \bar{y})}{\underbrace{S \sqrt{\frac{1}{n} + \frac{1}{m}}}} \sim S(t)_{n+m-2}$$

Se no le σ sono molto diverse \Rightarrow α α diventa test approssimato

Inoltre non è detto che siano NORMALI \Rightarrow è un'ip. che posso verificare con una certa precisione, ma di solito non posso dire che sono esatti.

Sulla distribuzione $S(t)$ posso calcolare l'int. di

$$\text{confidenza} \rightarrow \left\{ t_{\frac{\alpha-1}{2}}, t_{\frac{\alpha+1}{2}} \right\}$$

$$\text{ovvero } P\left(-t_{\frac{1-\alpha}{2}} \leq t \leq t_{\frac{1-\alpha}{2}}\right) = 1-\alpha$$

code simm. perché sto chiedendo $\mu_1 = \mu_2$
 se avessi per es. $\mu_1 > \mu_2$ allora
 avrei int. asimmm.

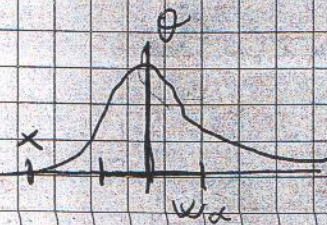
$$\Rightarrow \text{SOSTITUENDO: } P\left(\bar{x} - \bar{y} - t_{\frac{1-\alpha}{2}} S_p \leq \mu_x - \mu_y \leq \bar{x} - \bar{y} + t_{\frac{1-\alpha}{2}} S_p\right)$$

Spesso uso questi int. di conf. per risolvere un test d'ipotesi.

es. $X \sim F(x, \theta)$

$H_0: \theta = \theta'$

es \Rightarrow media



Voglio costruire $\overline{W}_{1-\alpha} \Rightarrow P(X \in \overline{W}_{1-\alpha} | H_0) = 1 - \alpha$
 = regione di accettazione (è funzione di θ)

misuro $x \rightarrow$ suppongo x' fuori da $\overline{W}_{1-\alpha}$

ritorno il test (serie infin.?) cambiando θ'

e stavolta x finisce sul bordo,

\Rightarrow si può definire, a partire da una serie di test di ip. un intervallo di valori di θ' accettabili:

$$I_x(\theta) = \{ \theta' : x \in \overline{W}_{1-\alpha} \}$$

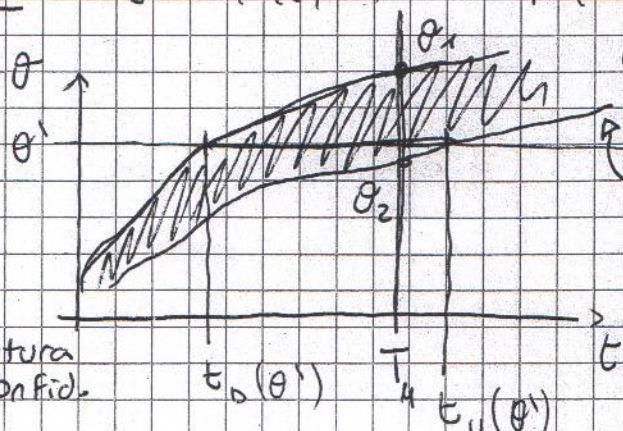
$$x \in \overline{W}_{1-\alpha} \Leftrightarrow \theta' \in I_x(\theta)$$

questo ha prob. $1-\alpha$ e quindi anche l'altro

$\Rightarrow I_x(\theta)$ ha il livello di confidenza $1-\alpha$ (?)

es. $t \sim F(t, \theta)$

$$P(t \in \overline{W}_{1-\alpha} | \theta') = \alpha$$



\hookrightarrow trovo t_L e t_U ma variano in funz. di θ'

queste 2 linee sono MONOTONAMENTE CRESCENTI

misuro $t_M \Rightarrow \{ \theta_2, \theta_1 \} = I_t(\theta)$

è come prima, rapp. graficamente

30/11/2011

Paragone di due campionamenti: test di Student.

$$X \sim N(\mu_x, \sigma^2)$$

$$Y \sim N(\mu_y, \sigma^2)$$

(varianze = e non note)

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{y} = \frac{\sum y_i}{m}$$

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$s_y^2 = \frac{\sum (y_i - \bar{y})^2}{m-1}$$

Se $n = m$ posso combinare i due campioni a dare un unico s^2 , altrimenti la stima migliore è:

$$s^2 = \frac{s_x^2(n-1) + s_y^2(m-1)}{n+m-1}$$

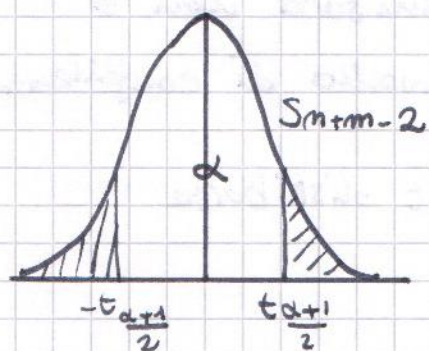
N.B.: possiamo farlo solo se non abbiamo il minimo sospetto che le σ possano essere diverse.

Test di Student:

$$t = \frac{\mu_x - \mu_y - (\bar{x} - \bar{y})}{\left(s \sqrt{\frac{1}{n} + \frac{1}{m}} \equiv S_p \right)}$$

Se non posso usare s^2 : $t = \frac{\dots}{\sqrt{\frac{s_x^2}{n^2} + \frac{s_y^2}{m^2}}} \dots$

$t \sim \mathcal{D}_{n+m-2}(t)$ ← solo se $x, y \sim N$. Se le varianze sono diverse, questo test è approssimato.



$$P\left(-\frac{t_{\alpha/2}}{2} \leq t \leq \frac{t_{\alpha/2}}{2}\right) \equiv \alpha$$

definisce l'intervallo

$$P\left(\bar{x} - \bar{y} - t_{\frac{1-\alpha}{2}} S_p \leq \mu_x - \mu_y \leq \dots + t_{\frac{1-\alpha}{2}} S_p\right) = \alpha$$

Otengo un intervallo di confidenza

a livello α .

C'è una definizione di intervallo di confidenza che si riferisce in modo esplicito al test di decisione (definizione di Neyman).

$$X \sim f(x, \theta) \quad H_0: \theta = \theta' \quad \bar{w}_\alpha$$

$$P(X \in \bar{w}_\alpha | H_0) = 1 - \alpha$$

Ci sono vari modi di costruire w_α , ma questo problema è simmetrico ($H_1: \theta \neq \theta'$)

\Rightarrow lo centro in θ' .

Supponiamo di misurare $x \in w_\alpha$, $x \notin \bar{w}_\alpha$

\Rightarrow l'ipotesi viene rigettata.

H_0 viene accettata se x si trova sul bordo, e

θ' viene messo nell'intervallo $d\theta$:

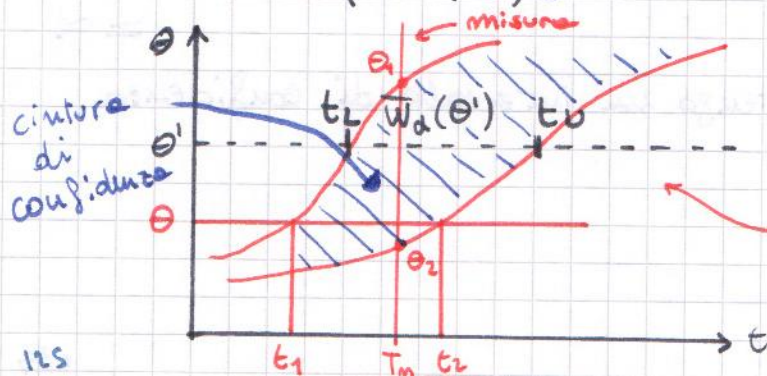
$$I_x(\theta) \equiv \{\theta: x \in \bar{w}_\alpha\}$$

NB.: $\bar{w}_\alpha = \bar{w}_\alpha(\theta')$. Per variare θ' costruisco altre ipotesi H_0 nell'intervallo in cui θ' è accettabile.

$$x \in \bar{w}_\alpha \iff \theta \in I_x(\theta)$$

I è funzione di $x \Rightarrow$ è un intervallo aleatorio (prima della misura). Dopo la misura non è più una probabilità ma un intervallo di confidenza.

Supponiamo di avere una statistica t distribuita come $t \sim f(t, \theta)$.



\bar{w}_α soddisfa:

$$P(t \in \bar{w}_\alpha(\theta')) = \alpha$$

queste curve sono monotonamente crescenti al variare di θ .

L'intervallo di confidenza può anche coprire il valore atteso per solo una piccola frazione. Ad es.: massa del neutrino \rightarrow intervallo sistematicamente negativo.

Supponiamo che il nostro intervallo di conf. denza sia costruito in modo che
$$\int_{-\infty}^{t_1} f(t, \theta) dt = \frac{1-\alpha}{2}$$

$\{\theta_1, \theta_2\} = I_t(\theta)$ $\alpha\%$ per costruzione.

In una frazione α dei casi l'intervallo "abbraccia" il valore vero θ . In questi casi T_n viene misurato in $\{t_1, t_2\}$.

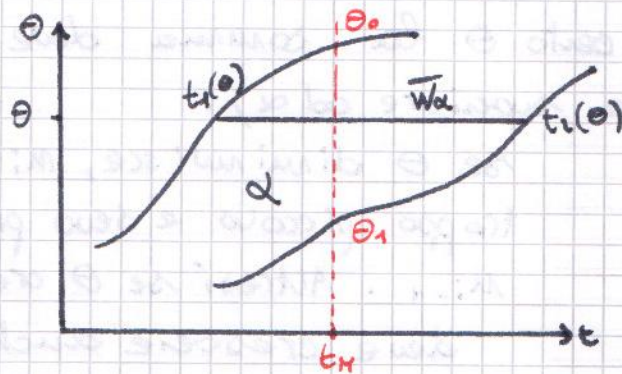
Questo è vero anche per intervalli non simmetrici o semirette \Rightarrow ottengo dei sup o degli inf.

1/12/2011

In certi casi possiamo fare misure e stimare θ da una statistica t :

$$x \sim f(x, \theta) \longrightarrow t \sim g(t, \theta)$$

Quando ciò non è possibile si definisce un intervallo di confidenza.



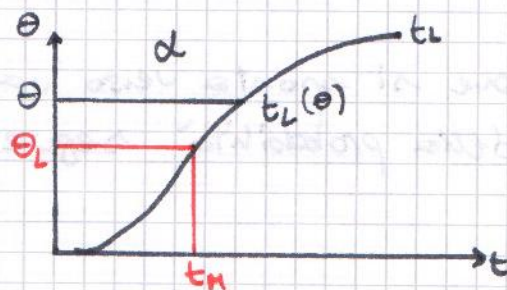
$$H_0: \theta = \theta_0$$

$$H_1: \theta \neq \theta_0$$

Ogni intervallo orizzontale ha contenuto di probabilità α .

Se l'intervallo vero è θ , la probabilità che quello che ricavo dalla misura ~~lo~~ lo ricopra è α .

$$t \in \bar{w}_\alpha \quad \theta \in I_t(\theta) = \{ \theta : x \in \bar{w}_\alpha(\theta) \}$$



Supponiamo di misurare t_n .

$$\theta_L = t_L^{-1}(t_n)$$

$$\text{Quindi: } t \leq t_L \Rightarrow$$

$$\theta_L \leq \theta$$

↓

$$P(t \leq t_L) = \alpha \Rightarrow P(\theta_L \leq \theta) = \alpha$$

(Le due affermazioni sono equivalenti).

Ho trovato un limite inferiore ai valori di θ .

Consideriamo un'esperienza di conteggio di una variabile **discreta** ad es. binomiale.

$$n \sim \text{Bin}(n; p, N)$$

In questo caso la definizione funzionale di $t_L(\theta)$ non si può fare: integrale \rightarrow somma, e la

somma non è continua.

Dobbiamo sostituire l'uguaglianza con una disuguaglianza. Si preferisce "sovracoprire" il valore di θ , per convenzione:

$$P(M \leq M_0(\theta)) = \alpha \longrightarrow P(M \leq M_0(\theta)) \geq \alpha$$
$$\Rightarrow M = \max n \text{ t. c. } P(\dots) \geq \alpha.$$

Supponiamo che per un certo θ la somma deve finire ad m_i (è appena superiore ad α):



se θ diminuisce, m_i diventa troppo piccolo e devo prendere m_{i-1} . Altrimenti se θ cresce deve crescere anche n \Rightarrow abbiamo una funzione a gradini.

Ci sono parecchi valori di θ per cui $M \leq n_i$...

$$\forall \theta \in A \quad P(M \leq m_i) \geq \alpha$$

$$\forall \theta \in B \quad P(M \leq m_{i+1}) \geq \alpha$$

$$* \theta = \sup A \Rightarrow P(M \leq m_i) = \alpha$$

Le probabilità diminuiscono al crescere di θ . Al sup di ogni intervallo, per

ottenere nuovamente una disuguaglianza, devo passare al gradino successivo.

Aumentando θ la distribuzione si sposta verso valori grandi di n . Il massimo della probabilità segue la distribuzione.

$$P(M \leq m_{i-1} | \theta_L) = \alpha$$

Se complemento questa relazione ottengo:

$$P(M \geq m_i | \theta_L) = 1 - \alpha$$

Così ottengo un intervallo che ha copertura nominale del parametro $\alpha\%$.

La probabilità che θ_L sia davvero un limite inferiore per tutti i possibili θ si dice **probabilità di contenimento**.

La probabilità di copertura è perfettamente definita nel caso continuo, non lo è nel caso discreto.

Prendiamo il più piccolo processo poissoniano possibile.

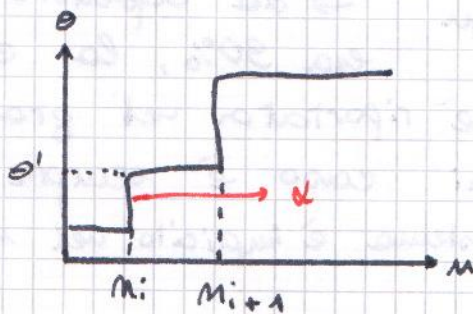
$$\mu = 0 \quad \alpha = 90\% \quad \Rightarrow \mu_{up} = 2.3$$

$$P(\mu < 2.3) = 1$$

Vogliamo trovare la condizione $P(m \geq m_0 | \theta) \geq \alpha$.
Anche in questo caso dobbiamo scegliere m_0 in base a θ , dato α .

Possiamo scegliere:

$$m_0 \equiv \begin{cases} \min m_0 : P(m \geq m_0 | \theta) \geq \alpha \\ \max m_0 : P(m \geq m_0 | \theta) \geq \alpha \end{cases}$$



Supponiamo di aver misurato m_i . Allora la relazione che dobbiamo considerare è:

$$P(m \geq m_{i+1} | \theta) = \alpha$$

$$(P(m \geq m_i | \theta) \geq \alpha)$$

Misurato $m_i \Rightarrow \theta_u = \theta_u$ t.c. $\begin{cases} P(m \geq m_{i+1} | \theta_u) = \alpha \\ P(m \leq m_i | \theta_u) = 1 - \alpha \end{cases}$
Questo definisce θ_u . (l'upper limit)

$$P(\theta \leq \theta_u) = \alpha$$

Nel caso del lower limit, $m \leq m_0 \Leftrightarrow \theta \geq \theta_l$ (nel caso continuo). Nel caso discreto l'implicazione non è doppia:

$$\{m \leq m_0\} \subseteq \{\theta \geq \theta_l\}$$

$$\alpha \leq P(m \leq m_0) \leq P(\theta \geq \theta_l)$$

Allo stesso modo, per l'upper limit:

$$\{m \geq m_0\} \subseteq \{\theta \leq \theta_u\}$$

$$\alpha \leq P(m \geq m_0) \leq P(\theta \leq \theta_u)$$

Torniamo alla **probabilità di copertura** e consideriamo il caso di un processo poissoniano su cui ho misurato

$$m = 0. \quad m \sim \frac{e^{-\mu} \mu^m}{m!}$$

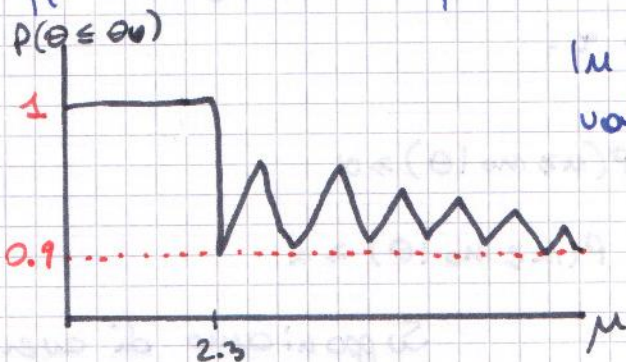
$$\Theta_\mu \rightarrow P(\mu=0|\Theta_\mu) = 1 - \alpha \quad \alpha = 90\%$$

$$e^{-\Theta_\mu} = 0.1$$

$$\Theta_\mu = 2.3$$

$$P(\mu \leq \Theta_\mu) = 90\% \quad \text{pu costruzione!!}$$

La probabilità di copertura in funzione di μ è:



In realtà 90% è un valore asintotico che si ottiene solo ad alti valori di μ .

\Rightarrow La copertura richiesta era 90%, la copertura

effettiva dipende da μ ed è riportata nel grafico: parte dal 100% e in ogni caso è sempre $\geq 90\%$
 \Rightarrow SOVRA COPERTURA. Questo problema è implicito nel metodo stesso.

Supponiamo di voler misurare un'efficienza $\hat{E} = \frac{m}{N}$ e di ottenere $m=N=10$: a questo punto l'intervallo di confidenza NON può superare 1, e con le formule sopra si ottiene $(0,741; 1)$.
E' per una Poissoniana, se abbiamo misurato $m=10$ e vogliamo $\alpha = 68.3\%$:

$$m=10 \quad [6.9, 14.3]$$

$$m=1 \quad [0.1, 3.3]$$

06/12/2011

Sebbene il metodo della massima verosimiglianza sia preferibile, il metodo piú utilizzato è il metodo del minimo χ^2 .

Si suppone che il minimo sia ottenibile con i migliori stimatori dei parametri veri.

I metodi dei minimi quadrati, della massima verosimiglianza e dei momenti forniscono risultati consistenti tra loro.

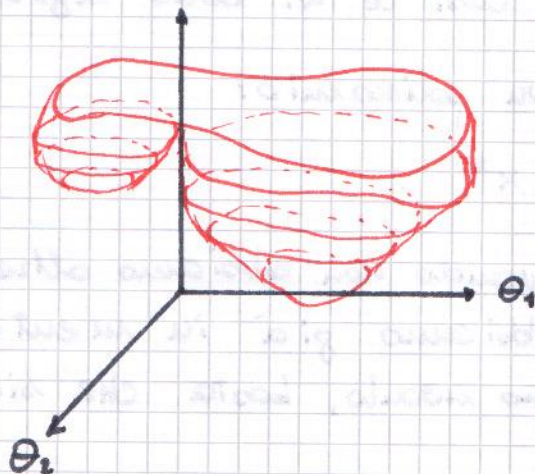
Metodo del mini-max \rightarrow si minimizza un solo valore: lo scarto massimo.

Nella pratica del metodo dei minimi quadrati ci si minimizza:

$$\chi^2 = \sum (\eta(x) - y(x))^2 w$$

con η valori "teorici" ed y dati sperimentali.

Non per forza η è lineare, sarà in generale una funzione $\eta(x, \theta)$. Essa può anche avere piú di un minimo:



Noi tratteremo solo il caso "semplice" in cui c'è un solo minimo e sappiamo già arrivare vicino ad esso. In questo caso il problema è lineare nei parametri θ :

$$\eta(x, \theta) = \sum \phi_i(x) \theta_i$$

Spesso è anche possibile linearizzare in modo diretto o con un'espansione di Taylor. Ad esempio:

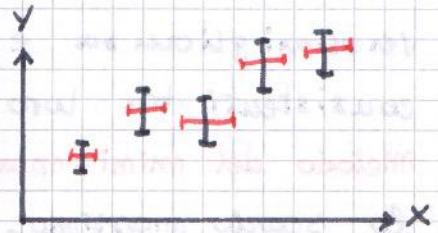
$$y = a e^{bx}$$

$$\log y = \log a + bx$$

Tratteremo anche il caso in cui le x non sono soggette ad errore.

$$\{x_i, y_i \pm \sigma_i\}$$

↑
supposte misurate
senza errore



Nel caso pratico questo non accade: il trattamento più di questo caso è molto complicato e non lo faremo.

x_i = variabili indipendenti.

y_i = funzioni degli x_i , soggette ad errore.

$$\sum_{i=1}^n [\eta(x_i, \theta) - y(x_i)]^2 w_i$$

w_i è il peso associato alle y_i .

Qualora le y siano distribuite normalmente, $w_i = \frac{1}{\sigma_i^2}$.

Nella maggior parte dei casi le w_i sono uguali $\forall i$.

Esempio di fitting con un polinomio:

$$\eta(x, \vec{\theta}) \equiv \theta_0 + \theta_1 x + \theta_2 x^2$$

Questo si usa soprattutto quando non abbiamo altre informazioni che i dati misurati. Se abbiamo già in mente un modello teorico possiamo usarlo, basta che sia lineare nei parametri.

Esempio: $\eta(x, \theta) = a + bx$

Chiamo y il vettore delle y_i : $y = \{y_i\}$ $x = \{x_i\}$

Individuo il modello η con una matrice di disegno:

$$\eta(x, \theta) = A \theta \quad \theta = \begin{Bmatrix} a \\ b \end{Bmatrix}$$

In questo caso,

$$A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \Rightarrow A \cdot \theta = \begin{bmatrix} a + b x_1 \\ \vdots \\ a + b x_m \end{bmatrix}$$

Dovrò poi paragonare i vettori $A \cdot \theta$ ed y , usando quindi $(y - A\theta)$.

Vediamo il caso piú semplice: $\sigma_i = \sigma \forall i$

$$X^2 = \sum [y_i - (a + b x_i)]^2 \quad \text{non metto i pesi, sono tutti uguali.}$$

$$\begin{cases} \frac{\partial X^2}{\partial a} = -2 \sum (y_i - a - b x_i) = 0 \\ \frac{\partial X^2}{\partial b} = -2 \sum (y_i - a - b x_i) x_i = 0 \end{cases}$$

Questo sistema è particolarmente semplice:

$$\begin{cases} \sum y_i - n a - b \sum x_i = 0 \\ \sum x_i y_i - a \sum x_i - b \sum x_i^2 = 0 \end{cases}$$

2 equazioni in 2 incognite.

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

Vediamo quindi la trattazione matriciale:

$$X^2 = (y - A\theta)^T (y - A\theta)$$

$$\frac{\partial}{\partial \theta} X^2 = -2 A^T (y - A\theta) = 0$$

$$\Rightarrow A^T y - A^T A \theta = 0$$

Moltiplicando l'espressione per $(A^T A)^{-1}$ otteniamo:

$$\hat{\theta} = (A^T A)^{-1} A^T y$$

Questo era il caso non pesato. Il più delle volte avremo dei pesi, anche se spesso tutti uguali.

N.B.: A non deve essere singolare. A meno che non sovradimensioniamo il problema, le matrici A non dovrebbero essere malcondizionate.

Se abbiamo molti parametri è possibile che gli errori commessi nel misurare le y condizionino completamente la soluzione: dovremo usare qualche metodo migliore. (\rightarrow vedi calcolo numerico \rightarrow sist. lineari).

$$\eta = E(y) = A\theta$$

θ vettore di L elementi $\Rightarrow A = n \times L$, n numero misure. Consideriamo il caso generale in cui le y sono correlate tra loro:

$$\text{Cov}(y) \equiv V_y$$

La forma quadratica allora si scrive:

$$X^2 = (y - A\theta)^T V_y^{-1} (y - A\theta)$$

N.B.: se le y_i sono tutte indipendenti, V_y è diagonale.

Per definizione: $V(y_i, y_j) = \frac{1}{n} \sum (y_i - \bar{y}_i)(y_j - \bar{y}_j)$

Supponiamo che nel nostro caso V_y sia diagonale o diagonalizzabile. Allora:

$$\frac{\partial X^2}{\partial \theta} = -2 A^T V_y^{-1} (y - A\theta) \equiv 0$$

e va risolto il sistema di L equazioni:

$$(A^T V_y^{-1} A) \theta = A^T V_y^{-1} y$$

che restituisce, una volta risolto,

$$\hat{\theta} = (A^T V_y^{-1} A)^{-1} A^T V_y^{-1} y$$

N.B.: Se vogliamo fare il calcolo componente per componente:

$$X^2 = \left(y_i - \sum_j A_{ij} \theta_j \right)^T (V_y^{-1})_i \left(y_m - \sum_n A_{mn} \theta_n \right)$$

e poi facciamo le derivate rispetto ad ogni θ_i .

$$\begin{aligned} E(\hat{\theta}) &= (A^T V_y^{-1} A)^{-1} A^T V_y^{-1} A E(y) \\ &= (A^T V_y^{-1} A)^{-1} A^T V_y^{-1} A \theta = \theta \end{aligned} \Rightarrow \hat{\theta} \text{ è NON DISTORTO}$$

Inoltre:

$$\text{Var } \hat{\theta} = S V_y S^T \quad S = \frac{\partial \hat{\theta}}{\partial y} = (A^T V_y^{-1} A)^{-1} A^T V_y^{-1} V_y V_y^{-1} (A^T V_y^{-1} A)^{-1}$$

La matrice di covarianza è autotrasposta: $(V_y^{-1})^T = V_y^{-1}$

$$\Rightarrow \text{Var } \hat{\theta} = (A^T V_y^{-1} A)^{-1}$$

C'è poi una terza proprietà, il **teorema di Gauss-Markov**. Possiamo costruire molti stimatori di θ lineari nelle y . Supponiamo di avere gli stimatori $\hat{\theta}$ e $\hat{\epsilon}$:

$$\hat{\theta} = (A^T A)^{-1} A^T y$$

$$\hat{\epsilon} = U y \quad : \quad E(\hat{\epsilon}) = \theta$$

$$E(Uy) = \theta \Rightarrow U A \theta = \theta \Rightarrow A U = \mathbf{1}_{L \times L}$$

Si può dimostrare che gli elementi diagonali della matrice di covarianza dei $\hat{\epsilon}$ sono tutti più grandi o uguali di quelli di $\hat{\theta}$.

Teorema di Gauss-Markov

Se \exists altro stimatore $\hat{\epsilon}$ non distorto, allora $\hat{\theta}$ è più preciso: $\text{Var}(\hat{\epsilon}) = U U^T \sigma^2$ (con $\sigma^2 = \text{Var}(\hat{\theta})$) ($V_y \equiv \sigma^2 \mathbf{1}$).

La dimostrazione è complicata. Provarne per esercizio utilizzandolo:

$$\begin{cases} A U = \mathbf{1} \\ U U^T \equiv (A^T A)^{-1} + (U - (A^T A)^{-1} A^T) (U - (A^T A)^{-1} A^T)^T \end{cases}$$

Se due matrici di covarianza degli stimatori sono uguali solo nel caso:

$$U = (ATA)^{-1} AT$$

ma allora in questo caso: $\hat{\theta} = (ATA)^{-1} AT y = \hat{\theta}$

⇒ La varianza degli stimatori ottenuti con il metodo dei minimi quadrati è minore di quella degli stimatori ottenuti con altri metodi.

Se definisco $X^2_{MIN} \equiv X^2 : \hat{\theta} = \theta$
allora

$$E(X^2_{MIN}) = E[(y - A\hat{\theta})^T V_y^{-1} (y - A\hat{\theta})] = n - L$$

Per impostare la dimostrazione utilizziamo l'approssimazione

$$V_y = \sigma^2 \mathbb{1}$$

Utilizziamo: $y - A\theta = y - A\hat{\theta} + A(\hat{\theta} - \theta)$

$$\begin{aligned} X^2 &= (y - A\theta)^T \frac{1}{\sigma^2} (y - A\theta) = \\ &= \frac{1}{\sigma^2} (y - A\hat{\theta} + A(\hat{\theta} - \theta))^T (y - A\hat{\theta} + A(\hat{\theta} - \theta)) = \\ &= \frac{1}{\sigma^2} [(y - A\hat{\theta})^T (y - A\hat{\theta}) + (\hat{\theta} - \theta)^T A^T A (\hat{\theta} - \theta)] \end{aligned}$$

I termini misti sono nulli:

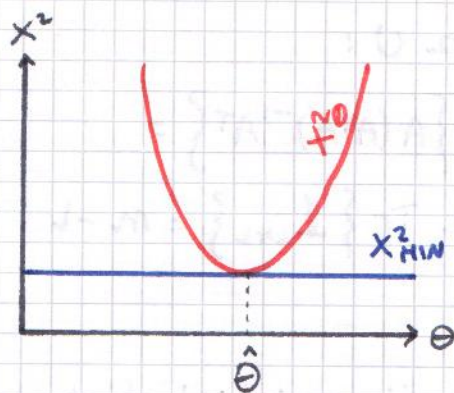
$$(y - A\hat{\theta})^T A (\hat{\theta} - \theta) \leftarrow \text{sostituire } \hat{\theta} \text{ e fare il calcolo} \rightarrow = 0$$

Chiamo:

$$X^2_{MIN} = \frac{1}{\sigma^2} (y - A\hat{\theta})^T (y - A\hat{\theta})$$

$$X^2_{\theta} = \frac{1}{\sigma^2} (\hat{\theta} - \theta)^T A^T A (\hat{\theta} - \theta)$$

Rimane: $X^2 = X^2_{MIN} + X^2_{\theta}$



X_0^2 ha un minimo in $\hat{\theta}$ ed è un paraboloide.

Saltando qualche passaggio, possiamo arrivare a scrivere:

$$X_{MIN}^2 = (y - A\theta)^T \left(\quad \right) (y - A\theta)$$

↑
matrice di covarianza

e arriveremo a scrivere la matrice di covarianza nella forma:

$$V_y = \mathbb{I}_{m \times m} - A(A^T A^{-1})^{-1} A^T$$

Supposto vero questo, allora X_{MIN}^2 è della forma:

$$X_{MIN}^2 = v^T U v$$

$$E(v) = 0$$

$$\text{Var}(v) = \sigma^2 \mathbb{I}_{m \times m}$$

Consideriamo la forma quadratica $Q \equiv y^T U y$.

$$E(y) = \eta \quad \text{Var}(y) = \sigma^2 \mathbb{I}_{m \times m}$$

Dimostriamo ora che $E(Q) = \sigma^2 \text{Tr}(U) + \eta^T U \eta$:

$$E(Q) = E\left(\sum_{j=1}^m (U_{jj} y_j y_j - U_{jj} \bar{y}_j \bar{y}_j) + \sum_{j=1}^m U_{jj} \bar{y}_j \bar{y}_j \right) =$$

$$= \sum_{j=1}^m U_{jj} (y_j y_j - \bar{y}_j \bar{y}_j) + \sum_{j=1}^m U_{jj} \bar{y}_j \bar{y}_j = \eta^T U \eta$$

$$= \sigma^2 \sum_{j=1}^m U_{jj} + \eta^T U \eta = \sigma^2 \text{Tr}(U) + \eta^T U \eta$$

$$* = \sum_{j=1}^m U_{jj} E(y_j y_j - \bar{y}_j \bar{y}_j) = \sum_{j=1}^m U_{jj} \sigma^2 \delta_{jj}$$

Per completare la dimostrazione basta calcolare la

traccia della matrice unitaria U :

$$\begin{aligned} \text{Tr}(\mathbb{I}_{m \times m} - A(A^T A)^{-1} A^T) &= m - \text{Tr}\{A(A^T A)^{-1} A^T\} = \\ &= m - \text{Tr}\{A^T A (A^T A)^{-1}\} = m - \text{Tr}\{\mathbb{I}_{L \times L}\} = m - L \end{aligned}$$

Dovremo ora considerare il caso di variabili distribuite normalmente, e trovare il modo di semplificare il calcolo dei θ : dal punto di vista della propagazione degli errori.

07/12/2011

$$\eta(x) = a + bx$$

$$A = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \quad A\theta = \begin{pmatrix} a + bx_1 \\ \vdots \\ a + bx_m \end{pmatrix}$$

$$\{x_i, y_i \pm \sigma_i\}$$

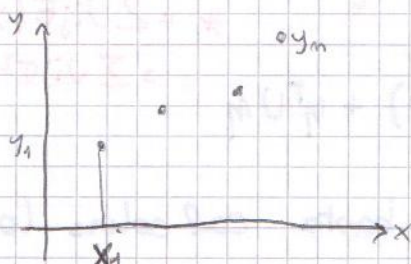
$$\theta = \{\theta_1 \dots \theta_m\}$$

Il metodo consiste nel minimizzare il termine $\chi^2 = (y - A\theta)^T W (y - A\theta)$ dove W è la matrice dei pesi.

La soluzione **ANALITICA** in questo caso è unica ed è data da $\hat{\theta} = (A^T W A)^{-1} A^T W y$.

- La soluzione è unica.
- $E(\hat{\theta}) = \theta$
- $\text{Var}(\hat{\theta}) = (A^T W A)^{-1}$
- Teor. di Gauss-Markov: ogni altro stimatore, lineare nei parametri, non distorto, ha una varianza non inferiore a $\hat{\theta}$.
- $E(\chi^2_{\text{min}}) = m - L$

Trattiamo ora il caso in cui la serie di punti sperimentali sia modellabile con una dipendenza lineare di y da x .



$$W = \frac{1}{\sigma^2} \mathbb{I}$$

La matrice dei pesi è l'inverso della matrice di covarianza delle y . Ora stiamo supponendo le y indipendenti e tutte con la stessa varianza.

$$\hat{\theta} = (A^T A)^{-1} A^T y$$

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & \dots & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = A^T A = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$A^T y = \begin{bmatrix} \sum x_i y_i \\ \sum y_i^2 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

$$\uparrow \det^{-1}(A^T A)$$

$$\hat{\theta} = (A^T A)^{-1} A^T y = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i \\ n \sum x_i y_i - \sum x_i \sum y_i \end{bmatrix}$$

Se faccio una piccola variazione ad \hat{a} , diciamo negativa, la retta che meglio interpola i miei punti dato il nuovo \hat{a} avrà un nuovo \hat{b} , cioè $\hat{b} + \delta \hat{b} > \hat{b}$.
 \rightarrow i miei parametri sono correlati.

Come posso cambiare il mio modello per evitare questa correlazione?

$$y(x) = a + bx \quad \rightarrow \quad \phi(x) = a + b(x - \bar{x})$$

$A^T A$ diventa una matrice diagonale. Senza troppa fatica posso farla diventare unitaria.

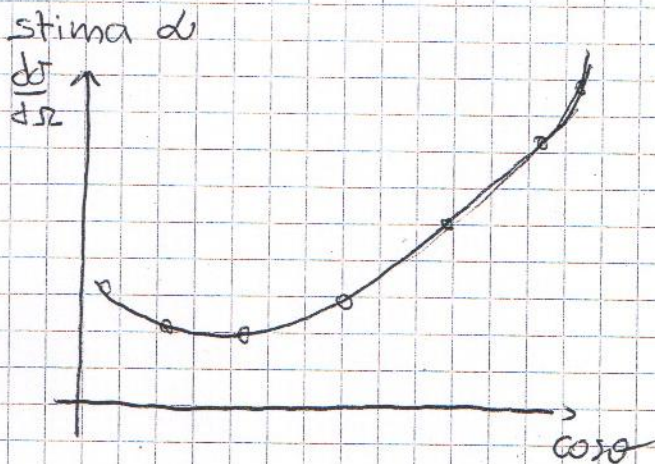
In realtà vedremo che conviene utilizzare come modelli dei polinomi ortogonoli ai dati. 140

Per L devo sapere la distrib. \Rightarrow devo definire il processo

χ^2 è più "generico".

es. $e^+e^- \rightarrow \mu^+\mu^-$ $\frac{d\sigma}{d\Omega} \propto 1 + \cos^2\theta + \underbrace{d\cos\theta}$

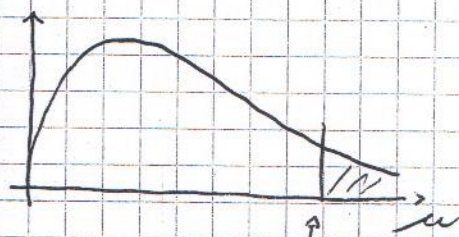
questo è dovuto
a 2 canali del
dec. (γ e Z^0)
 \Rightarrow interferiscono
[perché P è violata]



Calcolo il χ^2

$$\chi^2_{\min} \sim \chi^2_{n-1} \Rightarrow \hat{d}$$

Allora:



se tipo sono
qui

e χ^2_{\min} è grande (?)
che succede?

Se non avessi messo $d\cos\theta$
avrei trovato χ^2 brutto \Rightarrow test
di signif. mi avrebbe messo
nesso in evidenza un problema.
(o nella teoria, o nell'esperimento)

- QUANDO HO CURVA. MA NON DESCRITTA DA MODELLO

devo estrarne fuori interpolazione basandosi sui

soli dati \Rightarrow in questo caso va meglio il min χ^2 .

Es. devo ho un segnale + fondo

\Rightarrow per togliere il fondo



devo sapere come va SOTTO IL SEGNALE \Rightarrow si
interpola e poi si va a RICAVARE LA FONZ.
sotto il picco.

Fitto con polinomi ORTOGONALI \Rightarrow mi garantiscono
 max della stabilità numerica.

(Nota \Rightarrow la formul. $\hat{\theta}$ = matr. a -1 è va bene per
 descr. analitica, ma per i conti per i programmi
 usano strumenti più avanzati.

FITTO CON: $\eta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots$

$\phi(x) = \phi_0 \phi_0 + \phi_1 P_1(x) + \phi_2 P_2(x) + \dots$ (questa
 è la
 curva
 d'interp.)
 con $\sum_i P_e(x_i) P_m(x_i) = \delta_{em}$

$\underline{es} \rightarrow$ avevamo visto $P_0 \propto 1$
 $P_1 \propto x - \bar{x}$
 si costruiscono così:

$-2P_j(x) = x P_{j-1}(x) - \alpha P_{j-1}(x) - \beta P_{j-2}(x)$

$\alpha = \sum_i x_i P_{j-1}^2(x_i)$ $\beta = \sum_i x_i P_{j-1}(x_i) P_{j-2}(x_i)$

La α si arriva da: $-2 \sum_i P_e^2(x_i) = 1$

supp. caso con $\sigma^2 I = V_y$

$\hat{\theta} = \underbrace{(A^T A)^{-1}}_{\substack{\downarrow \\ I \\ L \times L \text{ mat. } L = \# \text{ par. } (?)}}$ $A^T y = A^T y$

si ottiene, con i
 pol. ortog.:

$\Rightarrow \hat{\theta}_j = \sum_i P_j(x) y_i$

\Rightarrow Ne segue che:

θ_i sono scorrelate

$V_{\theta_j} = \sigma^2 \sum P_j(x_i)^2 = \sigma^2$ $Cov(\hat{\theta}_e, \hat{\theta}_m) = \sigma^2 \delta_{em}$

"REMISSIONE LINEARE ?" "REGRESI. UN ?" come si chiama?

se ho un dato $x \rightarrow$ estrapolo $\phi(x)$

$\phi(x) = y^* = \sum_{j=0}^r P_j(x) \hat{\theta}_j$

Calcolo $\text{Var}(y^*)$ (la voglio piccola)

$$\text{Var}(y^*) = \sum P_j(x)^2 \text{Var}(\hat{\phi}_j) = \sigma^2 \sum P_j^2(x)$$

se non conosco σ^2 lo stimo dai dati:

$$X_{\min}^2 = \frac{1}{\sigma^2} (y - A\hat{\phi})^T (y - A\hat{\phi})$$

$$\hat{\phi} = A^T y$$

sostituisco

$$\rightarrow = \frac{1}{\sigma^2} (y^T y - \hat{\phi}^T \hat{\phi}) =$$

$$= \frac{1}{\sigma^2} \left(\sum_{i=1}^n y_i^2 - \sum_{j=0}^r \hat{\phi}_j^2 \right) = \frac{S_r}{\sigma^2}$$

grado del polinomio

=> da qui trovo

$$\hat{\sigma}^2 = \frac{S_r}{n-r}$$

=> Val di asp. di $\hat{\sigma}^2$ è σ^2 =>

questo è il val asp. di X^2_{\min}

=> HO TROVATO UN BUN
ESTIMATORE

Che succede se aumento

il grado del polinomio $r \rightarrow r+1$

$$\text{trovo } y_{r+1}^* = \sum_{i=0}^{r+1} P_i(x) \hat{\phi}_i = y_r^* + P_{r+1}(x) \hat{\phi}_{r+1}$$

questo è lo stesso di prima ho solo un termine in più.

La var. è:

$$S_r = \frac{1}{\sigma^2} (y^T y - \hat{\phi}^T \hat{\phi}) = S_r - \hat{\phi}_{r+1}^2$$

↳ quando il grado supera (?) (o r supera grado pol.)

=> ha interp. tutti i punti e oltre fu zero

$$\text{Ma } \text{Var}(y_{r+1}^*) = \text{Var}(y_r^*) + P_{r+1}^2(x) \sigma^2$$

Var aumenta \Rightarrow devo capire fino a che grado mi dà ~~meno~~ più info (aumento gde del polin., si adatta meglio ai punti) e dove fin dove invece mi aumenta solo il rumore.

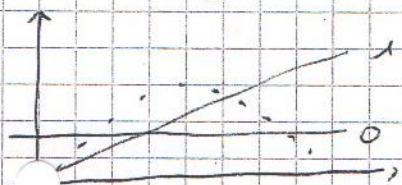
Se poi $y_i \sim N(\eta(x), \sigma^2)$

allora possiamo dire che: $\hat{\Phi}_j \sim N(\phi_j, \sigma^2)$

Se $\phi_j = 0$ (o lo metto = 0) $\rightarrow \hat{\Phi}_j$ sarà:

$$\frac{\hat{\Phi}_j}{\sigma} \sim N(0, 1) \quad \text{potrei farci test ip?}$$

Devo scegliere il grado giusto: es.



se è 0 o 1 non è va di sicuro.

\rightarrow grado r t.c. $\phi_r = \phi_{r+1} = \dots = \phi_{n-1} = 0$

prendo questa \uparrow (cioè quelli con grado $> r$)
come H_0 . non aggiungono info ??)

Cominciamo

Cominciamo calcolando i residui (S_r ?) per $n-1$

$$\rightarrow S_{n-1} = 0 = y^T y - \hat{\Phi}^T \hat{\Phi} = \sum_{i=1}^n y_i^2 - \sum_{i=0}^{n-1} \hat{\Phi}_i^2$$

$$\text{allora } S_r = \sum_{i=1}^n y_i^2 - \sum_{i=0}^r \hat{\Phi}_i^2 = \sum_{i=r+1}^n \hat{\Phi}_i^2 \Rightarrow \text{La statistica di } S_r \text{ è legata ai } \hat{\Phi} \text{ di grado } > r.$$

Se H_0 è vera (?)

$$\frac{\hat{\Phi}_r^2}{\sigma^2} \sim \chi_{r-1}^2 \rightarrow \hat{\sigma}_1^2 = \hat{\Phi}_r^2$$

$$\hat{\sigma}_r^2 = \frac{S_r}{n-r-1}$$
$$\frac{S_r}{\sigma^2} \sim \chi_{n-r-1}^2$$

2 estim. della var.

sono indep. perché

S_r dip. da $\hat{\Phi}_{r+1}, \dots$

\Rightarrow POSSO USARE TEST

DI FISHER PER

VEDERE SE STIMANO

LA STESSA VARIANZA

$$\textcircled{*} \rightarrow F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{S_{r+1} - S_r}{S_r} (n-r-1) \sim \textcircled{*}$$

$$\sim F(5)$$
$$1, n-r-1$$

Se questo mi dà ris. positivo

vuol dire che H_0 è vera ??

falso? (o invece ~~seg~~ è un passaggio successivo)

\rightarrow es. Vedi disperse.

Nota: se ho P-value alto meglio comunque

a vedere grado successivo, potrebbe esserci

Fluct. piccola che il pol. troppo basso non

vede. (trovo p-val. piccolo al grado succ.

\Rightarrow vuol dire che il grado succ. era necessario).

14/12/14

Abbiamo visto la regressione normale lineare.



$$f(\varphi, P) = \varphi_0 P_0 + \varphi_1 P_1(x)$$

$$\sum_{i=1}^m P_i P_j(x_i) = \delta_{ij}$$

$$P_0 = 1/\sqrt{n}$$

$$P_1(x) = \frac{x - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}}$$

$$\hat{\varphi} = A^T y$$

$$A = \begin{bmatrix} P_0 & P_1(x_1) \\ \vdots & \vdots \\ P_m & P_1(x_n) \end{bmatrix}$$

$$\hat{\varphi}_i = \sum_{j=1}^m y_j P_j(x_i)$$

$$\text{Var}(\hat{\varphi}_i) = \sigma^2$$

$$\eta(x) = \frac{\varphi_0}{\sqrt{n}} + \frac{x - \bar{x}}{s} \hat{\varphi}_1$$

$$s^2 = \sum (x_i - \bar{x})^2$$

Vorrei poi fare estrolazioni e/o interpolazioni:

$$\text{Var}(\eta) = \sigma^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{s^2} \right)$$

$$\frac{\eta(x) - y}{\sqrt{\text{Var}(\eta)}} \sim N(0, 1)$$

Se conosco σ^2 posso definire gli intervalli di confidenza:

$$\eta = \eta(x) \pm \lambda_\alpha \sigma \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{s^2}}$$



Se σ^2 non è noto, posso stimarlo con:

$$\hat{\sigma}^2 = \frac{X^2_{\min}}{n-2}$$

$$\frac{\eta(x) - y}{\sqrt{\text{Var}(\eta)}} \sim S_{n-2} \quad \eta(x) = \eta(x) \pm t_{\alpha} \sqrt{\hat{\sigma}^2} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S^2}}$$

Se σ^2 è noto, $X^2_{\min} \sim \chi^2_{n-2}$.

Mel caso del minimo X^2 si possono calcolare esattamente gli intervalli di confidenza, a differenza del caso della ML.

$$X^2(\theta) = (y - A\theta)^T V_y^{-1} (y - A\theta) + X^2_{\min} - X^2_{\min}$$

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

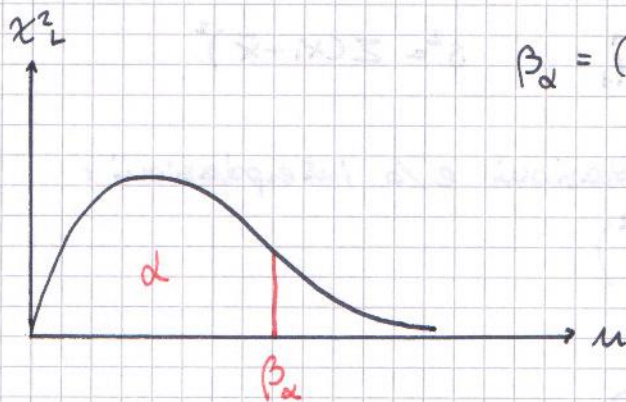
solo 3 χ^2 indipendenti.

l'numero dei parametri.

con $V_{\theta}^{-1} = A^T V_y^{-1} A$ Posso scrivere:

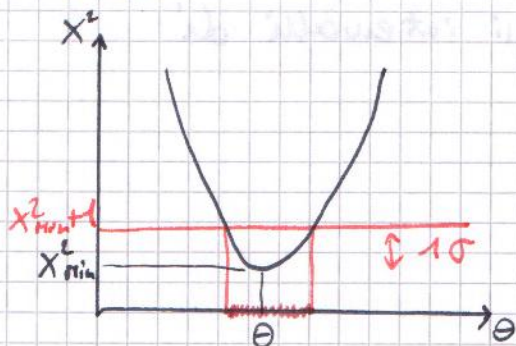
$$u = X^2(\theta) - X^2_{\min} = (\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta}) \sim \chi^2_L$$

e posso calcolare $P(u \leq \beta_{\alpha}) = \alpha$.



$$\beta_{\alpha} = (\theta - \hat{\theta})^T V_{\theta}^{-1} (\theta - \hat{\theta})$$

Consideriamo il caso $L=1$:



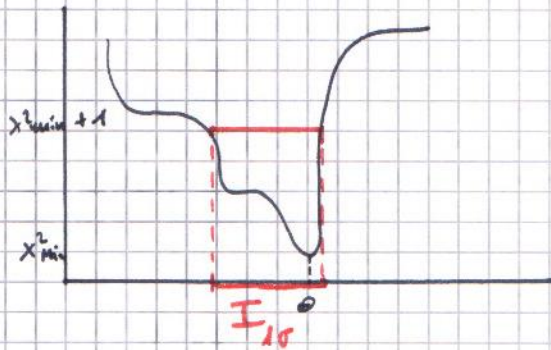
$$\frac{(\theta - \hat{\theta})^2}{\sigma_{\theta}^2} = P_{\alpha}$$

$$d = 69.3\%$$

$$\frac{(\theta - \hat{\theta})^2}{\sigma_{\theta}^2} = 1$$

Se il χ^2 non è lineare non è infrequente trovare X^2 asimmetrici - Saldando di $n\sigma$ definiamo comunque

Intervallo di confidenza a $1-\alpha$.



Con $L = 2$ abbiamo 2 parametri θ_1 e θ_2 . Supponiamo che siano scorrelati.

$$\text{Var}(\theta) = \begin{pmatrix} \sigma_{\theta_1}^2 & 0 \\ 0 & \sigma_{\theta_2}^2 \end{pmatrix}$$

$$\frac{(\theta_1 - \hat{\theta}_1)^2}{\sigma_{\theta_1}^2} + \frac{(\theta_2 - \hat{\theta}_2)^2}{\sigma_{\theta_2}^2} = \beta_\alpha$$

$$\alpha = 68.3\% \Leftrightarrow \beta_\alpha = 2.25$$

ufficio prof. stanza 136

