# Stima dei parametri

Saverio Vicario – saverio.vicario@ba.itb.cnr.it

# Definizioni

- Diremo che n v.a. $X_1$ ,..., $X_n$ costituiscono un campione casuale della variabile X se esse sono indipendenti e identicamente distribuite con la stessa legge X
- Si chiama statistica qualunque v.a. T funzione del campione ovvero T=t(X1,…, Xn)
- Dato un campione X1,…, Xn di variabili aleatorie con legge dipdente dal parametro θ chiameremo stimatore del parametro θla statistica T.
- T e' uno stimatore non distorto di θ se E(T)=θ
- T e' uno stimatore consistente se T converge a θ al crescere di n

# Esempi di stimatori

- Quindi uno stimatore è una qualunque funzione del campione. Un esempio di stimatore del parametro $\theta = \mu$ è la media aritmetica. Esso e consistente e non distorto. La media campionaria puo' essere usata come stima del
- La varianza campionaria S2C è un esempio di stimatore della varianza. La LGN afferma che esso è uno stimatore consistente della varianza; esso è comunque distorto e per tale motivo richiede una piccola correzione.

# Test di ipotesi

Saverio Vicario – saverio.vicario@ba.itb.cnr.it

# Test di ipotesi

- In questa lezione affronteremo il problema del test di ipotesi e vedremo che l’esito di un test è una decisione tra ipotesi alternative.
- Nella lezione sulla teoria delle probabilità abbiamo avuto un primo approccio alla valutazione di ipotesi alternative in ambito a bayesiano.
- Affronteremo qui in dettaglio l’approccio frequentistica e lo contrasteremo con l’approccio bayesiano

# Esempio

- Supponiamo di voler stabilire se una moneta è truccata oppure equa (non truccata). A tale scopo effettuiamo $n=100$ lanci della moneta (esperimento) ed osserviamo 38 volte testa (T). Da questo risultato deduciamo che la probabilità empirica di ottenere T in un lancio è $p=38/100$. Possiamo affermare che la moneta è truccata oppure dobbiamo attribuire al caso il fatto di aver osservato 38 T in 100 lanci di una moneta non truccata?

# Definizione del problema

- Analizziamo gli effetti del caso. Supponiamo che la moneta sia equa, ovvero $p = ½$. In questa ipotesi il numero di T ($N_T$) in $n=100$ lanci è una v.a. Binomiale $B(100, ½)$. Ricorda che $E(N_T)=np=50$. Osserva che la deviazione rispetto alla media dei nostri dati sperimentali è $50 – 38 = 12$.
- Qual è la probabilità, nel caso di moneta equa, di ottenere risultati sperimentali uguali o più estremi di quelli osservati?
- Si noti la dizione "più estremi", non semplicemente una differenza positiva più grande di 12
- Il numero di interesse e' il valore assoluto $|50 – 38| = 12$

# Svolgimento

- Assumendo l'ipotesi H: X=B(100,0.5)
- L'evento di cui uno vuole conoscere la probabilità sotto l'ipotesi di interesse ovvero la P(D|H) si può definire come segue:
- P{ x≤50-12  U x≤50+12}
- Px≤38  U x≥62}
- F(x≤38) + ( 1-F(x≤61))
- Dunque traducendo in comandi R:

```
> pbinom(38,100,0.5)+(1-pbinom(61,100,0.5))
```

# Interpretazione

- P(D|H)= 0.02097874
- Nell’ipotesi di moneta equa solo due volte su 100 osserverò risultati uguali o peggiori di quelli osservati e quindi sono portato a concludere che la moneta è truccata.
- Ma la soglia di rischio che uno e’ disposto a correre varia
- E’ necessaria dunque aggiungere qualche definizione e formalizzare il processo decisionale

# Formalizzazione dell'esempio

- Iniziamo con l'osservare che ci sono due possibili ipotesi alternative:
- H0: la moneta è equa;
- H1: la moneta è truccata.
- A ciascuna ipotesi, in generale, associamo il valore di qualche parametro rilevante per il problema in oggetto. In questo caso il parametro potrebbe essere $N_T$(numero di T in 100 lanci) oppure la probabilità p di ottenere T in un lancio. Nel primo caso assoceremo $N_T= 50$ ad H0e $N_T \neq 50$ ad H1, nel secondo assoceremo $p= ½$ad $H_0$, e $p \neq ½$ ad $H_1$.

# Impostazione di un test

- Dal tipo di ipotesi scelte e' possibile calcolar e il P(D|H) solo sull'ipotesi delimitata p=1/2 l'altra essendo difficile da calcolare con i metodi sviluppati fin'ora ( ma in realtà sarebbe difficile in ogni caso).
- Il tipico inquadramento frequentista definisce l'insieme delle ipotesi possibili (es. p $\in$R)
- Un ipotesi zero su un sottoinsieme fisso delle ipotesi possibili (es. p =1/2)
- Un ipotesi 1 o alternativa che e' rappresenta $\overline{H_0}$

ovvero l'insieme di tutti i valori che non soddisfano H0

Se P(D|H0) scende al di sotto di un certo valore H0 diventa non credibile e dunque H1 viene accettato

# Impostazione di un test

- In generale lo sperimentatore sceglie come ipotesi nulla quella più conservativa e utilizza i dati sperimentali per verificare se è possibile rigettare l’ipotesi nulla in favore di quella alternativa di cui e’ realmente interessato.
- O meglio tra le ipotesi incluse nell’ipotesi alternativa c’e’ un ipotesi che vuole distinguere dall’ipotesi zero

- Ad esempio, nello studio dell'espressione di un gene X in una patologia Y, supponiamo che il livello di espressione di X è uguale in soggetti sani (H) ed in soggetti affetti dalla patologia (D) (ipotesi nulla). Dopo, misuriamo il livello di espressione di X in un gruppo di soggetti sani $X_H$ ed in un gruppo di soggetti malati $X_D$(dati sperimentali). Se $X_H$-$X_D$ è grande allora decido di rigettare H0in favore dell'ipotesi che X si esprime in modo diverso tra sani e malati (ipotesi alternativa).

# Nota

- Nota che le decisioni prese sulla base di test statistici sono solo probabilisticamente corrette. Avremo sempre la possibilità di commettere degli errori ed è importante calcolare la probabilità di commettere tali errori. Vedremo il legame tra il valore della soglia α e la probabilitàdi errore. Nel nostro esempio, se il numero di T in 100 lanci è 80 allora sono portato a dire che H1 è vera. Tale decisione è solo probabilisticamente corretta in quanto anche nell'ipotesi in cui H0 è vera posso osservare 80 T in 100 lanci.

# Tipi di errori

in generale possiamo commettere due tipi di errori:

- affermare che H0 è falsa quando H0 è vera;
- affermare che H0 è vera quando H0 è falsa.
- È importante controllare le probabilità di commettere tali errori.
- Nel nostro esempio, tali probabilità dipendono dalla differenza δ. Se δ è grande allora $P\{|p-½| > \delta\}$ è piccola se H0 è vera ($p=½$) e quindi la probabilità di commettere un errore del primo tipo è piccola. Ma aver scelto un δ grande ha come conseguenza quella di affermare spesso che H0 è vera anche quando essa è falsa, e quindi di aumentare la probabilità di commettere un errore del secondo tipo. Quindi il valore di δ deve essere scelto opportunamente al rischio che si vuole correre

# Tipi di errore

- Chiameremo errore di prima specie quello che si commette rigettando H0 quando H0 è vera.
- Chiameremo errore di seconda specie quello che si commette accettando H0 quando H0 è falsa.
- Il primo tipo di errore e' direttamente stimato da P(D|H0) e dunque e' perfettamente stimabile
- Il secondo tipo di errore generalmente non e' stimabile visto che l'ipotesi alternativa include un largo ventaglio di ipotesi puntali.
- E' bene dunque associare un rischio maggiore nel caso degli errori di primo tipo che sono sempre conoscibili piuttosto che in errori in quelli di secondo tipo

# Definizioni

- **livello del test**: la soglia α di probabilita di errore ( o rischio) tipo 1 che siamo disposti a prendere (tipici $\alpha=0.05$ o $\alpha=0.01$)
- **regione critica**: l'evento δ che conduce al rifiuto dell'ipotesi nulla H0 assumendo un certo livello α del test.
- **potenza del test** : $P(D_\alpha|H1^*)$, dove H1 e i dati sono fissati su dei valori predefiniti, con $D_\alpha$ dati generati assumendo H1 vera. Da una misura di probabilità di rifiutare H0 assumendo che H0 sia effettivamente falsa.
- **Livello β**: stima dell'errore di secondo tipo. 1-potenza del test
- **Significatività del test o p-value:** il valore di P(D|H0) o probabilità di errore di primo tipo.
-

# Un buon test

- Un test è buono quando α e p-value sono piccoli e la potenza è grande.
- α piccolo garantisce che gli errori di prima specie tollerati sono piccoli
- p-value piccolo (piu' piccolo di α) indica che l'ipotesi nulla e' stata rigettata
- La potenza del test e' grande allora vuol dire che se l'ipotesi alternativa

# Approssimazione continua del test

- Per illustrare meglio le varie componenti di un test, svolgeremo il test utilizzando le approssimazioni consentite dal Teorema del Limite Centrale.
- L'insieme di conteggi di teste T può essere visto oltre che come un esperimento Binomiale con dimensione 100 anche come 100 campioni di una distribuzione di Bernoulli
- Il valore medio dei successi seguirà una legge Normale con media e varianza presa dalla distribuzione di di Bernoulli.
- $E(N_T)=np \quad Var(N_T)=np(1-p)$
- $E(X)=p \qquad var(X)=p(1-p)/n \quad \rightarrow X: N(p,p(1-p)/n)$

# Impostazione

- Usando come statistica p e non $N_T$
- NT=30 n=100
- Allora X medio uguale p=0.38
- H0=0.5 e H1≠0.5
- $\delta=|0.5-0.38|=0.12$

# Svolgimento con approssimazione Normale

- Assumendo l'ipotesi H: X=100 B(1,0.5)
- X~N(p,np(1-p)/2)
- L'evento di cui uno vuole conoscere la probabilità sotto l'ipotesi di interesse ovvero la P(D|H) si può definire come segue:
- Evento D =x≤0.50-0.12  U x≤50+12
- P( x≤0.50-0.12  U x≤50+12)
- P(x≤0.38 ) + P(x≥0.62}
- F(x≤0.38) + ( 1-F(x≤o.61))
- Dunque traducendo in comandi R:
- Mean=0.5; Sd=(0.5*(1-0.5)/100)^0.5

```
> pnorm(0.38,mean=Mean,sd=Sd)+(1-pnorm(0.61,
  mean=Mean,sd=Sd))
> pbinom(38,100,0.5)+(1-pbinom(61,100,0.5))
```

# Geografia dei termini di un test

Eseguiamo i seguenti comandi su R.

```
delta=qnorm(0.05/2,mean=0.5, sd=(0.5*(1-0.5)/100)^0.5)
plot(0:100/100,dnorm(0:100/100,mean=0.38, sd=(0.38*(1-0.38)/100)^0.5),type='l',lty=2)
polygon(c(0.2,20:40/100,delta,delta),c(0,dnorm(c(20:40/100,delta),mean=0.38, sd=(0.38*(1-0.38)/
    100)^0.5),0) ,col=2, border=NA)
polygon(c(delta,delta,40:100/100,1),c(0,dnorm(c(delta,40:100/100),mean=0.38, sd=(0.38*(1-0.38)/
    100)^0.5),0) ,col=3, border=NA)
lines(0:100/100,dnorm(0:100/100,mean=0.5, sd=(0.5*(1-0.5)/100)^0.5))
polygon(c(0.2,20:40/100,delta,delta),c(0,dnorm(c(20:40/100,delta),mean=0.5, sd=(0.5*(1-0.5)/
    100)^0.5),0) ,col=4, border=NA)
delta2=qnorm(1-0.05/2,mean=0.5, sd=(0.5*(1-0.5)/100)^0.5)
polygon(c(delta2,delta2,60:100/100,1),c(0,dnorm(c(delta,60:100/100),mean=0.5, sd=(0.5*(1-0.5)/
    100)^0.5),0) ,col=4, border=NA)
abline(h=0)
points(0.38,0)
legend('topleft',legend=c('Potenza del test','Errore II tipo o beta','Errore I tipo o alpha'), fill=2:4)
legend('topright',legend=c('Ipotesi nulla o H0','ipotesi alternativa o H1'), lty=c(1,2))
legend('right', legend=c('Osservato'), pch=1)
```

# stime della potenza per un campione di 100

```
Mean=0:50/100;Sd=(Mean*(1-Mean)/
   100)^0.5;plot(0:50/100,pnorm(qnorm(0.05/2,0.5,
   (0.5*0.5/100)^0.5 ), Mean,Sd), type='l')
> title(main=c('Potenza del test di livello 0.05 su H0=0.5
   valutata con H1 da 0 a 0.5', ' con campione n=100'  ) )
>Mean=0.38 ;Sd=(Mean*(1-Mean)/
   1:100)^0.5;plot(1:100,pnorm(qnorm(0.05/2,0.5,
   (0.5*0.5/1:100)^0.5 ), Mean,Sd), type='l')
title(main=c('Potenza del test di livello 0.05 su ipotesi H0
   0.5 contro H1=0.38', ' con campione n da 1 a 100'  ) )
```

# Dimensioni del campione

```
>Mean=0.5 ;Sd=(Mean*(1-Mean)/1:100)^0.5
>plot(1:100,pnorm(0.38, Mean,Sd), type='l')
> abline(h=0.05/2)
>title(main=c('Significativita del test su ipotesi H0 0.5
   contro H1=0.38', ' con campione n da 1 a 100'  )
>lines(1:100,pnorm(0.30, Mean,Sd), col=2)
```

Oppure fissando il campione a 100

```
>Sd=(Mean*(1-Mean)/100)^0.5
> plot(1:50/100,pnorm(1:50/100, Mean,Sd), type='l')
> abline(h=0.05/2)
```

# Disegno di un esperimento

- Dobbiamo valutare se la distribuzione del sesso dei nascituri in una certa popolazione soggetta a particolare effetti ambientali segue una proporzione di 0.5 considerando successo per F ( femmina).
- Stabiliamo che vorremmo distinguere effetti che possano spostare la distribuzione di $\delta$=0.05
- Quanti nascituri dobbiamo visitare per ottenere un test con livello di significanza superiore alla soglia stabilita di 0.05

# Svolgimento

- X~ N(p,p(1-p)/n)
- H0 p=0.5
- H1 p≠0.5
- H1* p=0.55 U p=0.45
- $\alpha$ =0.05
- Mean=0.5 ;Sd=(Mean*(1-Mean)/1:1000)^0.5
- min(which(pnorm(0.45, Mean,Sd)<0.05/2))
- Essendo la distribuzione normale simmetrica non serve controllare l'altro evento
- min(which(pnorm(0.55, Mean,Sd)>(1-0.05/2)))

# Spiegazione

- Esploro tutti le dimensioni di campionamento da 1 a 1000 e per ognuno di questi calcolo media e varianza
- Mean=0.5 ;Sd=(Mean*(1-Mean)/1:1000)^0.5
- Calcolo la probabilità di osservare la proporzione di interesse sotto l'ipotesi nulla per tutti i possibili campionamenti
- X=pnorm(0.45, Mean,Sd)
- Definisco chi ha probabilita' al di sotto della soglia predefinita
- Y=X>(1-0.05/2)
- Selezioni quelli soddisfano la relazione
- Z=which(Y)
- Prendo il valore minimo
- min(Z)

# Disegno esperimento (cont.)

- Voglio stimare la probabilità di ottenere valori significativi se credo che l'effetto sia una diminuzione della proporzione a 0.45 avendo stabilito di usare 385 pazienti
- Allora devo stimare P(D|H1*) ovvero la potenza del test

# Svolgimento

Calcolo il delta per la soglia 0.05 sotto l'ipotesi H0

- Mean=0.5 ;Sd=(Mean*(1-Mean)/385)^0.5
- Delta=qnorm(0.05/2, Mean, Sd)

Calcolo la probabilità di osservare il valore Delta sotto la specifica sotto ipotesi H1*

- Mean=0.45 ;Sd=(Mean*(1-Mean)/385)^0.5
- pnorm(Delta, Mean, Sd)
- Potenza =0.50
- l'esperimento ha una probabilita' del 50% di non dare risultati significativi se l'ipotesi H1* e' vera

# Svolgimento (cont.)

- Mettendo il campione a 1300 pazienti si puo’ avere una probabilità del 95% di avere risultati significativi
- Mean=0.5 ;Sd=(Mean*(1-Mean)/1300)^0.5
- Delta=qnorm(0.05/2, Mean, Sd)
- Mean=0.45 ;Sd=(Mean*(1-Mean)/1300)^0.5
- pnorm(Delta, Mean, Sd)

# Paragone con un approccio Bayesiano

- Notare come conoscendo la distribuzione di riferimento nel test Bayesiano e' possibile calcolare direttamente tutte le probabilità di interesse mentre nel caso frequentista solo il p-value e' perfettamente stimabile.
- Non svolgo il test della moneta perche con un approccio bayesiano richiede distribuzioni (dirichlet) che non fanno parte del programma. Comunque e' ancora risolvibile analiticamente.
- Per problemi più complessi le difficoltà dell'approccio Bayesiano impongono soluzioni "numeriche" approssimate e spesso mettono in imbarazzo il ricercatore nel definire una corretta probabilità a priori dei parametri

# Paragone con un approccio Bayesiano

- Un altro limite importate nell'interpretazione di un test frequentista e' che l'unico risultato utile e' il rifiutare l'ipotesi nulla.
- Se il p-value ovvero P(D|H0) e' grande non possiamo escludere che altre ipotesi possano spiegare ugualmente bene i dati (es.p=0.44, vedi grafico potenza del test)
- Se il p-value e' piccolo possiamo rifiutare H0
- In approccio Bayesiano le due o più alternative sono tutte esplorabili. Le ipotesi sono intercambiabili. Quando un ipotesi non e' puntuale e' possibile integrare le probabilità usando la probabilità a priori

# Limiti frequentisti in un applicazione

- In campo filogenetico spesso si vuole controllare se la divergenza fra sequenze di organismi e' regolare ed e' possibile assumere un "orologio molecolare" in cui la divergenza e' proporzionale al tempo.
- In un approccio frequentista si può calcolare la probabilità di che assumendo questa regola vera si possa osservare un certo insieme di dati, mentre non si possono calcolare le probabilità nell'infinita serie di ipotesi che fanno divergere le sequenze in maniera irregolare
- Dunque H0 deve essere definito come ipotesi orologio molecolare ma il ricercatore e' interessato a corroborarla e non rifiutarla.

# Test Bayesiano/Test Frequentista

Risultati del test diagnostico

Vero stato del paziente

| | **Positive** | **Negative** |
|---|---|---|
| D | TP | FN |
| H | FP | TN |

Se imponiamo H0=D e H1=H allora:

TP (TruePositive): soggetti positivi al test che sono malati.

FN (False Negative): soggetti negativi al test che sono malati. E' equivalente al Livello di significatività

FP (False Positive): soggetti positivi al test che sono sani. E' equivalente a $\beta$ o errore di secondo tipo

TN (TrueNegative): soggetti negativi al test che sono sani.

## Esempio

- La probabilità di recupero in soggetti affetti da cancro della cervice uterina è elevata in caso di individuazione precoce. Il Paptest è una procedura di screening che può individuare il cancro anche in casi asintomatici. Studi condotti negli anni 1972-1978 in 306 laboratori in 44 stati hanno rilevato che:
- $P(N \mid D) = 0.1625$ (*probabilità di un FN*);
- $P(P \mid D) = 0.8375$ (*Se*).
- $P(P \mid H) = 0.1864$ (*probabilità di un FP*);
- $P(N \mid H) = 0.8136$ (*Sp*).

# Valore predittivo di un test positivo

- Determiniamo $P(D \mid P)$
- *Applichiamo il teorema di Bayes:*
- $P(D \mid P) = P(P \mid D)\, P(D) / (P(P \mid D)\, P(D) + P(P \mid H)\, P(H))$
- *Sapendo che le donne affette da cancro della cervice sono 8.3 su 100.000 negli anni '82-'83:*
- $P(D) = 0.000083$ e $P(H) = 1 - P(D) = 0.999917$.
- *Allora:* $P(D \mid P) = 0.000373$.
- $P(D \mid P)$ *è detto valore predittivo di un test positivo.*

# Valore predittivo di un test negativo

- Determiniamo $P(H \mid N)$
- *Applichiamo il teorema di Bayes:*
- $P(H \mid N) = P(N \mid H)\, P(H/( P(N \mid D)\, P(D) + P(N \mid H)\, P(H))$

  *Nelle stesse ipotesi precedenti si ha:* $P(H \mid N) = 0.999983$.
- $P(H \mid N)$ *è detto valore predittivo di un test negativo.*
- *Questo indica che su 1.000.000 di donne risultate negative al Paptest, 999983 erano sane.*