UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di scienze economiche, aziendali, matematiche e statistiche "Bruno de Finetti"

Introduzione: inferenza e verosimiglianza

(Statistica c.p.)

Francesco Pauli

DEAMS

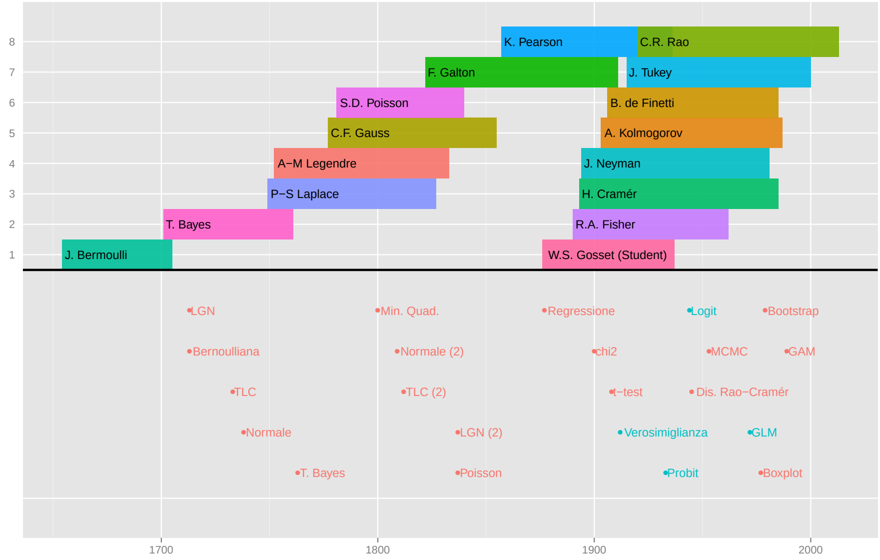
Università di Trieste

A.A. 2016/2017

Traccia del corso

1. Modelli statistici parametrici.
2. Inferenza con la verosimiglianza: in pratica.
3. Esemplificazione dell'inferenza basata sulla verosimiglianza per il modello lineare.
4. Inferenza con la verosimiglianza: alcuni aspetti teorici (proprietà degli SMV e quantità collegate, consistenza, distribuzione asintotica, riparametrizzazioni, aspetti computazionali).
5. Test basati sulla verosimiglianza.
6. La verosimiglianza profilo.
7. Principi di riduzione dei dati, statistiche sufficienti, sufficienti minimali e ancillari; principi di condizionamento e di verosimiglianza.
8. Aspetti pratici: studio della verosimiglianza con l'impiego del software statistico R: funzioni per la rappresentazione grafica e per l'ottimizzazione.

Un quadro storico



Indice

L'inferenza statistica

Inferenza

Inferenza è la produzione di una proposizione come conseguenza necessaria di una o più proposizioni. (De Morgan, Sillabo, 141-142)

- Traggo conclusioni certe a partire da premesse certe, ricade nel dominio della logica.
- Nell'inferenza statistica si sostituisce certe con incerte, sia per quanto attiene alle conclusioni che alle premesse.
- L'inferenza statistica consiste cioè nel trarre delle conclusioni, che saranno incerte, a partire da dei dati che sono soggetti a variazioni casuali.

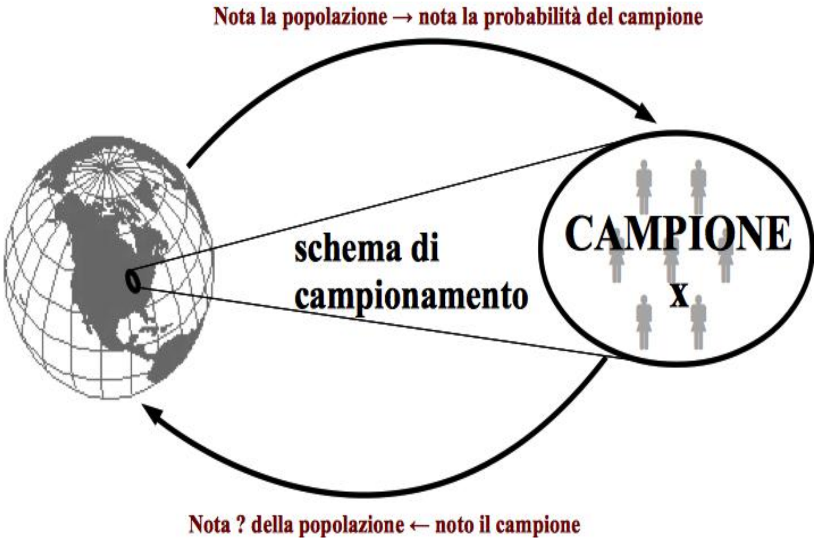
Esempio: popolazione finita

- Consideriamo la **popolazione** italiana maschile, e vogliamo determinare quanti individui posseggono due cromosomi y (cosiddetta sindrome 47-XYY).
- Non potendo osservare l'intera popolazione se ne seleziona un sottoinsieme casualmente e si osserva questo, il **campione**.
- Degli individui del campione si osservano i cromosomi e quindi quanti posseggono la terna XYY, sulla base di questa osservazione si traggono delle conclusioni sulla popolazione.
- Il processo di selezione casuale del campione rende casuali i dati, le conclusioni sulla popolazione saranno dunque casuali.

Inferenza statistica

L'inferenza statistica è quella parte della statistica che si occupa di indurre le caratteristiche di un aggregato dall'osservazione di una parte di esso.

Schema dell'inferenza



Indice

Prototipo di problema statistico

- **Popolazione**:
  - urna con N =10 palline;
  - di cui una proporzione θ è rossa, le altre sono bianche.
  - Quante sono le rosse?
- **Spazio parametrico**:

$$\Theta = \{0/10, 1/10, 2/10, \ldots, 9/10, 1\}$$

  (11 possibili valori.)
- Non possiamo osservare l'urna, ma possiamo osservare un
- **campione**: estraiamo con reimbussolamento n =5 palline e osserviamo il numero di palline rosse R.
- Spazio campionario: $\mathcal{R} = \{0, 1, \ldots, 5\}$

Meccanismo generatore e deduzioni

Noto θ, so che

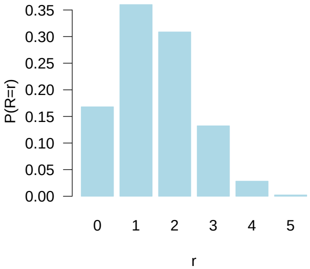
$$P(R = r) = \binom{n}{r}\theta^r(1-\theta)^{n-r}, \quad r = 0, \dots, n,$$

ad esempio se θ =0.3, l'urna contiene 3 palline rosse e 7 bianche, allora

$$P(R = 3) = 0.1323.$$

Insomma, posso ottenere la distribuzione di probabilità di R

r	Pr
0	0.1681
1	0.3601
2	0.3087
3	0.1323
4	0.0284
5	0.0024



La soluzione con le 'probabilità inverse'

- Supponiamo che l'urna sia stata riempita con un meccanismo casuale tale per cui
$$P(\theta = i/10) = p_i.$$
- Dato il campione $R = r$ useremmo allora il teorema di Bayes e scriveremmo, posto $\theta_i = i/10$,
$$P(\theta = \theta_i | R = r) = \frac{P(R = r|\theta = \theta_i)p_i}{\sum_{j=0}^{10} P(R = r|\theta = \theta_j)p_j}.$$
- Queste probabilità riassumerebbero l'informazione su θ dato $R = r$.
- Notiamo che, per poter arrivare a una distribuzione di probabilità su θ dobbiamo:
  - interpretare θ come v.a. (che questo sia un passo non da poco forse non è immediato),
  - avere una distribuzione iniziale, che viene da ...?
- Questo approccio va sotto il nome di statistica bayesiana.

Il modello

Ragioniamo invece in un altro contesto

- θ non è una v.a.
- di conseguenza, non ha senso, non esiste, $P(\theta = \theta_i)$.

Possiamo calcolare la probabilità dei possibili campioni per ogni possibile θ, ottenendo il modello

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	1	0.5905	0.3277	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000	0
1	0	0.3280	0.4096	0.3601	0.2592	0.1562	0.0768	0.0284	0.0064	0.0004	0
2	0	0.0729	0.2048	0.3087	0.3456	0.3125	0.2304	0.1323	0.0512	0.0081	0
3	0	0.0081	0.0512	0.1323	0.2304	0.3125	0.3456	0.3087	0.2048	0.0729	0
4	0	0.0005	0.0064	0.0284	0.0768	0.1562	0.2592	0.3601	0.4096	0.3280	0
5	0	0.0000	0.0003	0.0024	0.0102	0.0312	0.0778	0.1681	0.3277	0.5905	1

- Cioè il modello è l'insieme dei possibili meccanismi generatori dei dati tra cui "peschiamo" (qui, 11, le colonne della tabella).
- Come lo usiamo per inferire su θ?

Inferenza

Fissiamo il campione: sia $R = 3$

Notiamo che, se fosse $\theta = 0.2$, la probabilità di avere $R = 3$ sarebbe molto minore di quanto non sarebbe se fosse $\theta = 0.8$

$$P(R = 3; \theta = 0.2) = 0.0512 < 0.2048 = P(R = 3; \theta = 0.8)$$

attrae allora l'idea, tra i due valori di θ, di preferire quello che rende più probabile il campione effettivamente osservato.

Funzione di verosimiglianza I

Estendiamo questo ragionamento e usiamo la riga della tabella corrispondente a $R = 3$ come una serie di punteggi che permettono di creare un ordine di preferenza tra i diversi valori del parametro (e, più raramente, di escludere alcuni valori).

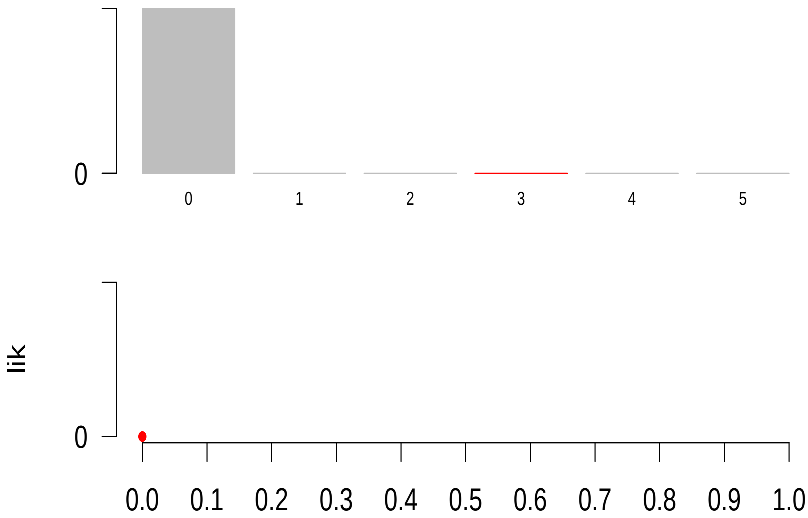
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	1	0.5905	0.3277	0.1681	0.0778	0.0312	0.0102	0.0024	0.0003	0.0000	0
1	0	0.3280	0.4096	0.3601	0.2592	0.1562	0.0768	0.0284	0.0064	0.0004	0
2	0	0.0729	0.2048	0.3087	0.3456	0.3125	0.2304	0.1323	0.0512	0.0081	0
3	0	0.0081	0.0512	0.1323	0.2304	0.3125	0.3456	0.3087	0.2048	0.0729	0
4	0	0.0005	0.0064	0.0284	0.0768	0.1562	0.2592	0.3601	0.4096	0.3280	0
5	0	0.0000	0.0003	0.0024	0.0102	0.0312	0.0778	0.1681	0.3277	0.5905	1

Ne discende allora che il valore che rende in assoluto più probabile il campione $R = 3$ è ...

Funzione di verosimiglianza II

- Questi ‘punteggi’ sono chiamati **verosimiglianze**.
- È importante il cambio di termine, che distingue l’approccio totalmente probabilistico (bayesiano) da quello, cosiddetto ‘classico’ che
    - interpreta θ come numero e non come v.a.,
    - non richiede una distribuzione di probabilità su θ.
- Selezionare il valore che rende più probabile il campione va sotto il nome di metodo della **massima verosimiglianza**.

Costruiamo la verosimiglianza



Rapporti di verosimiglianza

- Possiamo andare oltre la scelta del miglior valore.
- Confrontando i valori $L(\theta_i)$ possiamo stabilire quanto gli altri valori di θ sono meno attendibili
- ad esempio, $\theta_1 = 0.6$ è

$$\frac{L(0.6)}{L(0.5)} = \frac{\binom{5}{3}0.6^3 0.4^2}{\binom{5}{3}0.5^3 0.5^2} = 1.10592$$

  volte più plausibile di $\theta_2 = 0.5$.
- d'altra parte $\theta_1 = 0.6$ è 'ben'

$$\frac{L(0.6)}{L(0.9)} = \frac{\binom{5}{3}0.6^3 0.4^2}{\binom{5}{3}0.9^3 0.1^2} = 4.74074$$

  volte più plausibile di $\theta_3 = 0.9$.

Rapporti di verosimiglianza e verosimiglianza normalizzata

- Di particolare importanza saranno i rapporti rispetto al massimo valore.

$$\frac{L(\theta)}{\max L(\theta)}$$

(Nota: è il reciproco di quanto sopra.)

- A partire dalla verosimiglianza

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
3	0	0.0081	0.0512	0.1323	0.2304	0.3125	0.3456	0.3087	0.2048	0.0729	0

- Si ottiene, dividendo per il massimo, la cosiddetta verosimiglianza normalizzata

	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
3	0	0.0234	0.1481	0.3828	0.6667	0.9042	1.0000	0.8932	0.5926	0.2109	0

- Posso selezionare un gruppo di valori per θ (intervallo di confidenza) scegliendo una soglia e individuando i valori di θ per cui questo rapporto è superiore alla soglia stessa.

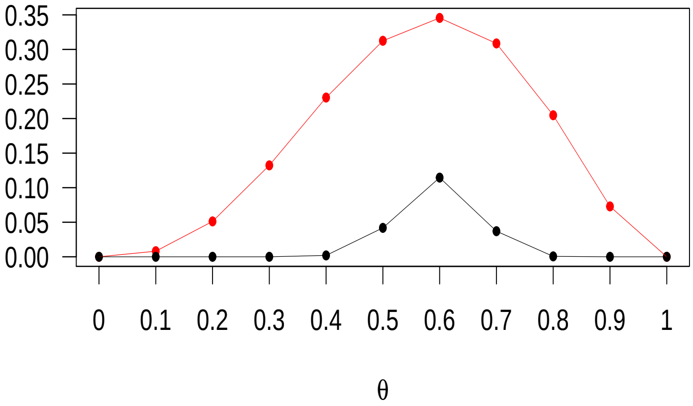
Rapporti e numerosità campionaria I

Cosa succede ai rapporti se aumentano le osservazioni?

- Supponiamo di osservare 50 palline, di cui 30 rosse.
- Il valore più probabile è sempre 0.6
- $\theta_1 = 0.6$ è
$$\frac{L(0.6)}{L(0.5)} = \frac{\binom{50}{30} 0.6^{30} 0.4^{20}}{\binom{50}{30} 0.5^{30} 0.5^{20}} = 2.73676$$
volte più plausibile di $\theta_2 = 0.5$.
- e $\theta_1 = 0.6$ è
$$\frac{L(0.6)}{L(0.9)} = \frac{\binom{50}{30} 0.6^{30} 0.4^{20}}{\binom{50}{30} 0.9^{30} 0.1^{20}} = 5.7340576 \times 10^6$$
volte più plausibile di $\theta_3 = 0.9$.

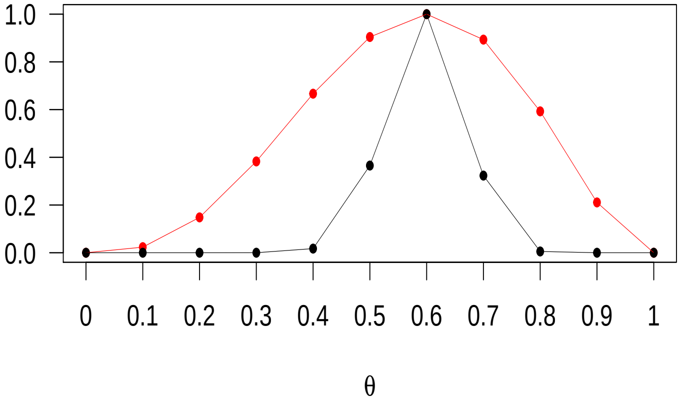
Verosimiglianza e numerosità campionaria I

Verosimiglianze per un campione di 5 palline di cui 3 rosse e per un campione di 50 palline di cui 30 rosse.

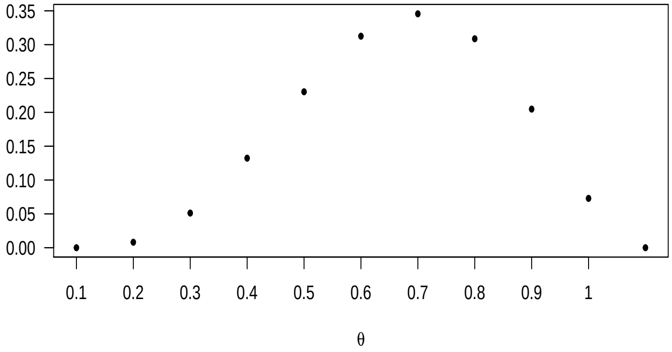


Verosimiglianza e numerosità campionaria II

Verosimiglianze normalizzate per un campione di 5 palline di cui 3 rosse e per un campione di 50 palline di cui 30 rosse.

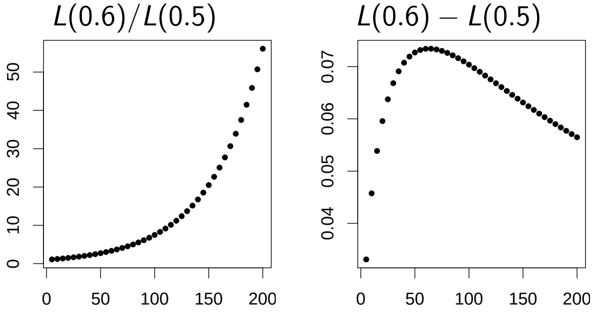


Verosimiglianza e numerosità campionaria III



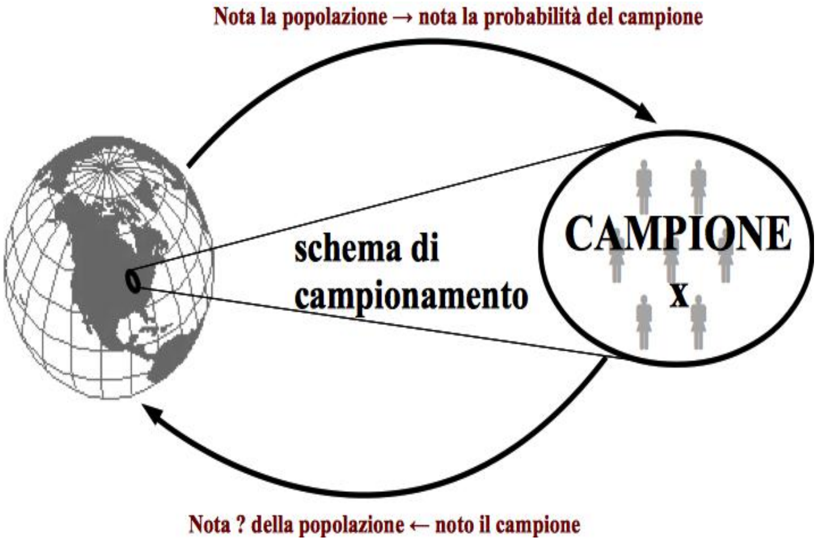
Perché i rapporti e non, ad es., le differenze? I

Campione	$L(0.6)/L(0.5)$	$L(0.6) - L(0.5)$
3 su 5	1.10592	0.0331
30 su 50	2.73676	0.0727
300 su 500	2.3570631×10^4	0.0364



Indice

Schema dell'inferenza

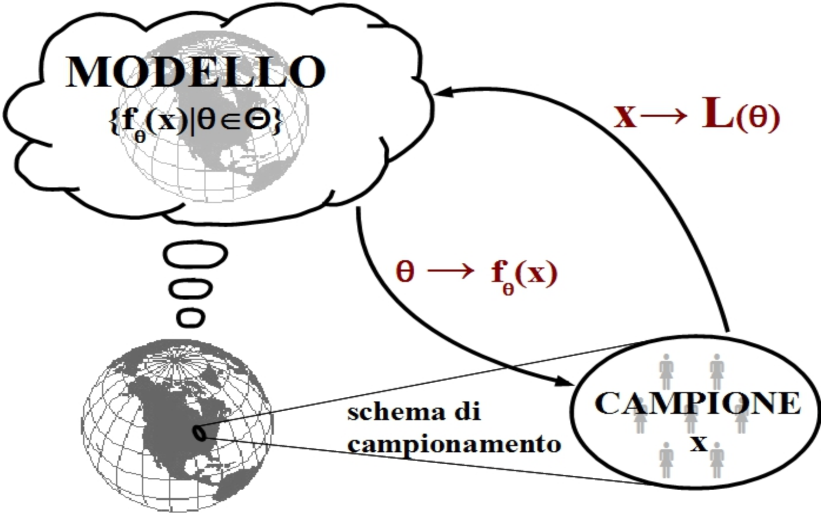


Schema dell'inferenza con modello

In termini formali

- Specifichiamo il *modello parametrico*, cioè un insieme di funzioni di probabilità del campione $\{f(x;\theta), \theta \in \Theta\}$.
- Fissato il campione x calcoliamo $f(x;\theta)$ (o $f_\theta(x)$) per ogni valore di θ.
- Interpretiamo il fatto che $f_{\theta_1}(x) > f_{\theta_2}(x)$ osservando che il campione x osservato è più probabile se il meccanismo generatore dei dati è θ_1 che non se è θ_2 stabilendo così un ordine di preferenza (vicinanza alle osservazioni) dei modelli.
- (La verosimiglianza fornisce in realtà anche un rapporto di preferibilità (nel senso che potremmo dire che, se $f_{\theta_1}(x) = 0.1$, $f_{\theta_2}(x) = 0.2$, $f_{\theta_3}(x) = 0.3$ potremo dire che θ_3 è tre volte più plausibile di θ_1, θ_2 è due volte più plausibile di θ_1).)

Schema dell'inferenza con modello



Indice

Un esempio con dei dati I

Consideriamo l'esempio seguente

- Ci chiediamo qual è la probabilità, che indichiamo con θ, che un nascituro sia maschio (con riferimento alla popolazione italiana).
- La popolazione è virtuale, i potenziali nascituri.
- Il campione è costituito da n nati, ad esempio i nati in un'area geografica in un certo periodo.
- Sia X il numero di maschi nel campione.
- X è una v.a. Binomiale di parametri n e θ

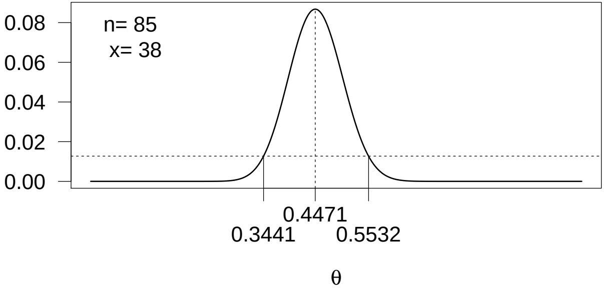
$$P(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

  per $x = 0, 1, \ldots, n$.
- Lo spazio parametrico è $\Theta = [0, 1]$.

Un esempio con dei dati II

- Nel 2010 a Muggia sono nati 38 maschi e 47 femmine.
- il campione conta cioè $n = 85$ nati e $x = 38$ maschi.
- Calcoliamo, per i possibili θ

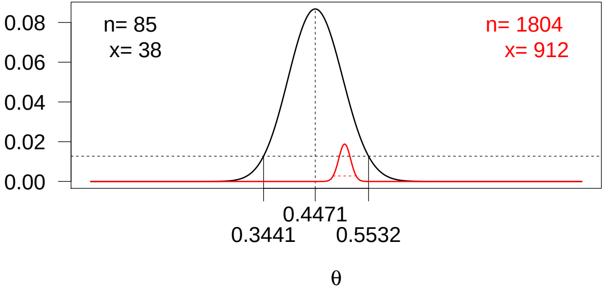
$$P(X = 38) = \binom{85}{38}\theta^{38}(1-\theta)^{47}$$



Un esempio con dei dati III

- Consideriamo un campione più ampio, la provincia di Trieste, dove su $n = 1804$ nati si sono avuti $x = 912$ maschi.
- Calcoliamo, per i possibili θ

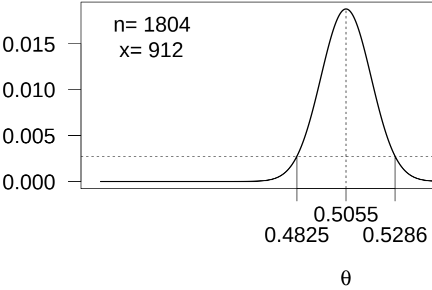
$$P(X = 912) = \binom{1804}{912}\theta^{912}(1-\theta)^{892}$$



Un esempio con dei dati IV

- Consideriamo un campione più ampio, la provincia di Trieste, dove su $n = 1804$ nati si sono avuti $x = 912$ maschi.
- Calcoliamo, per i possibili θ

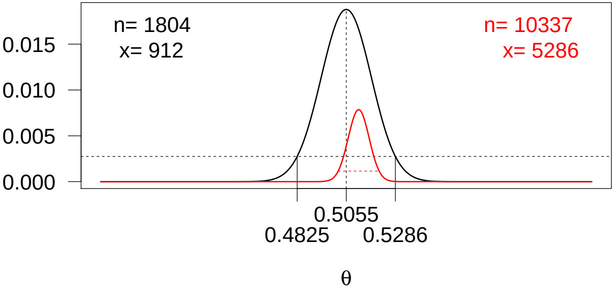
$$P(X = 912) = \binom{1804}{912} \theta^{912} (1-\theta)^{892}$$



Un esempio con dei dati V

- Ampliamo ancora il campione alla regione FVG, dove su $n = 10337$ nati si sono avuti $x = 5286$ maschi.
- Calcoliamo, per i possibili θ

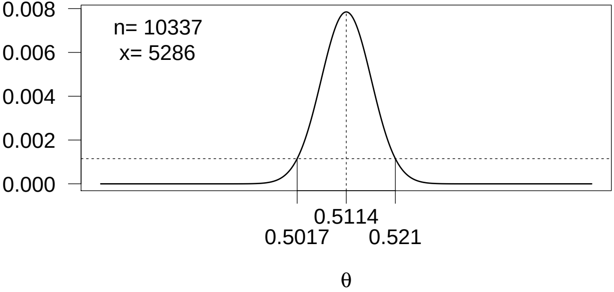
$$P(X = 5286) = \binom{10337}{5286} \theta^{5286} (1-\theta)^{5051}$$



Un esempio con dei dati VI

- Ampliamo ancora il campione alla regione FVG, dove su $n = 10337$ nati si sono avuti $x = 5286$ maschi.
- Calcoliamo, per i possibili θ

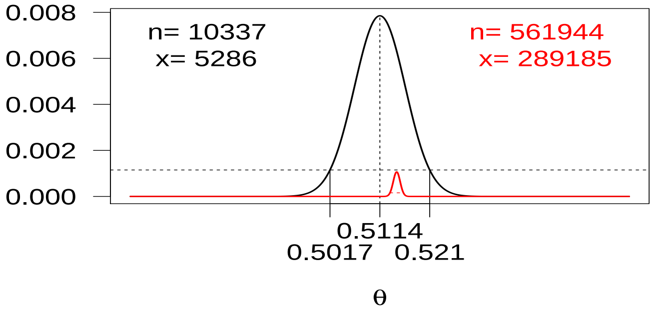
$$P(X = 5286) = \binom{10337}{5286} \theta^{5286} (1 - \theta)^{5051}$$



Un esempio con dei dati VII

- Infine, usiamo i dati nazionali, dove su $n = 561944$ nati si sono avuti $x = 289185$ maschi.
- Calcoliamo, per i possibili θ

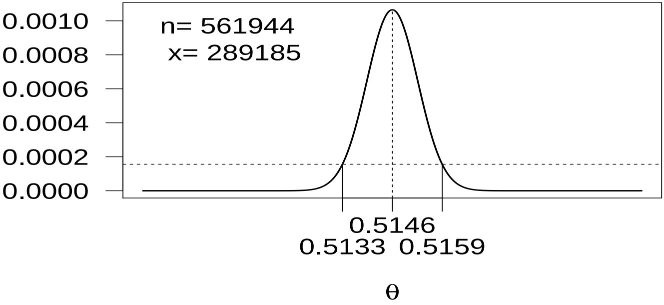
$$P(X = 289185) = \binom{561944}{289185} \theta^{289185} (1-\theta)^{272759}$$



Un esempio con dei dati VIII

- Infine, usiamo i dati nazionali, dove su $n = 561944$ nati si sono avuti $x = 289185$ maschi.
- Calcoliamo, per i possibili θ

$$P(X = 289185) = \binom{561944}{289185} \theta^{289185} (1 - \theta)^{272759}$$



Rapporti

Vogliamo valutare l'evidenza sperimentale pro o contro l'ipotesi $\theta = 0.5$, calcoliamo i rapporti

	maschi	femmine	totale	$\hat{\theta}$	$L(0.5)/L(\hat{\theta})$
Muggia	38	47	85	0.44706	0.62041610
provincia di Trieste	912	892	1804	0.50554	0.89505777
FVG	5286	5051	10337	0.51137	0.06915121
Italia	289185	272759	561944	0.51462	0.00000000

Trattamento analitico

formalizzazione del problema

- Campione X da una Binomiale di parametro θ e dimensione n

$$P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

- Per l'inferenza su θ si considera

$$L(\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

x è fissato!

Trattamento analitico

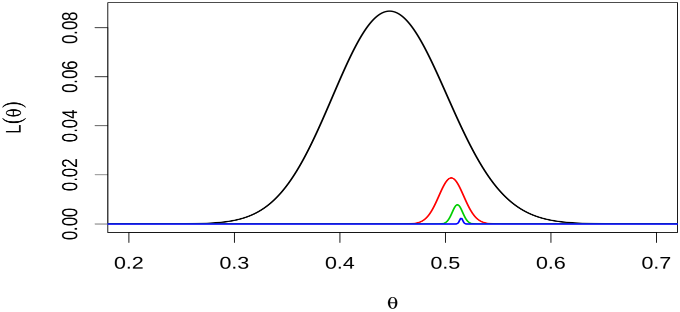
Il valore più plausibile

- Qual è il valore più plausibile?
- $\hat{\theta}$ tale che $L(\hat{\theta}) > L(\theta)$ per ogni θ.
- Cerco il massimo di $L(\theta)$ in $[0,1]$, che è uguale a cercare il massimo, in $[0,1]$, di
$$l(\theta) = x \log \theta + (n-x) \log(1-\theta)$$
- Calcolo la derivata prima e la eguaglio a 0
$$l_*(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$$
$$l_*(\theta) = 0 \Leftrightarrow \theta = x/n$$
- essendo la derivata seconda sempre minore di 0, questo è un massimo
$$l''(\theta) = -\frac{x}{\theta^2} - \frac{n-x}{(1-\theta)^2} < 0$$

Alcune rappresentazioni

Verosimiglianze

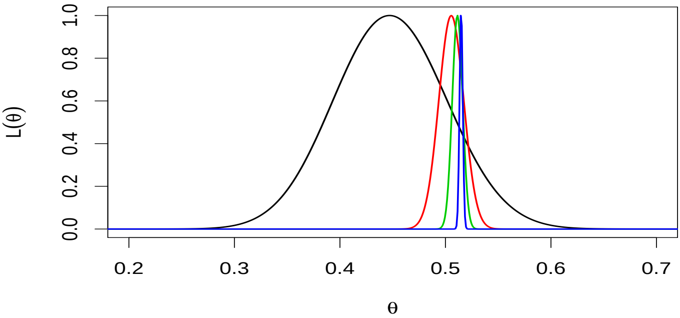
$$L(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$



Alcune rappresentazioni

Verosimiglianze normalizzate

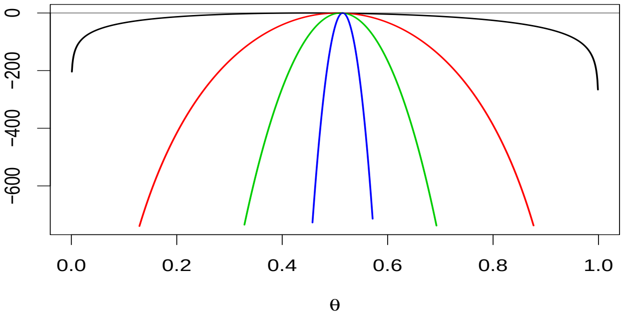
$$\frac{L(\theta)}{\max L(\theta)}$$



Alcune rappresentazioni

Log-verosimiglianze normalizzate

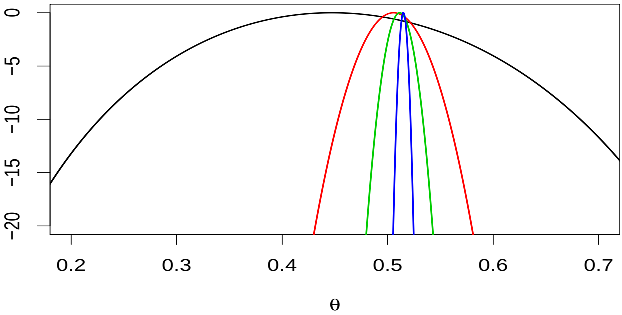
$$\log \frac{L(\theta)}{\max L(\theta)} = l(\theta) - \max l(\theta)$$



Alcune rappresentazioni

Log-verosimiglianze normalizzate

$$\log \frac{L(\theta)}{\max L(\theta)} = l(\theta) - \max l(\theta)$$



Indice

Schema dell'inferenza

Torniamo allo schema

Fasi e modi dell'inferenza statistica (parametrica) I

1. **Specificazione**: scelta di una famiglia di distribuzioni di probabilità, indicizzati da un parametro $\theta \in \Theta \subset \mathbb{R}^d$.
    - Correttamente specificata se contiene la 'vera' legge di probabilità del campione.
    - La "vera" legge di probabilità è una comoda idealizzazione,
    - (1) non è detto esista un vero modello per il meccanismo generatore dei dati (v. fenomeni sociali);
    - (2) non interessa davvero il vero modello, interessa un modello 'utile', cioè in grado di rispondere a ben specifiche domande.
    - cioè "tutti i modelli sono falsi, ma qualcuno è utile" (Cox?)
    - Quindi, il modello migliore può cambiare a seconda dello scopo.

Fasi e modi dell'inferenza statistica (parametrica) II

2. **Stima**: in questa fase l'obiettivo è discernere tra i possibili modelli (alias valori del parametro) quali siano più plausibili come meccanismo generatore dei dati.
   - *stima puntuale*: tiriamo a indovinare un valore per θ (*best guess*);
   - *stima intervallare*: cerchiamo di determinare un intervallo che pensiamo contenga il valore di θ;
   - *verifica d'ipotesi*: abbiamo un'ipotesi specifica su θ, $\theta > 0$ ad esempio e ci chiediamo se i dati più la smentiscano o più la confermino;
   - *previsione*: vogliamo prevedere valori futuri di x (la cui distribuzione di probabilità dipende da θ);
   - *decisioni*: dove a legge di probabilità delle conseguenze dipende da θ.

Fasi e modi dell'inferenza statistica (parametrica) III

3. **critica (validazione) del modello**: una volta stimato il modello, si cerca di capire se questo è effettivamente compatibile con i dati.
    - confronti dei quantili teorici versus quantili empirici;
    - validazione incrociata;
    - ...

Paradigmi dell'inferenza: la verosimiglianza è UN modo

- Come si fa inferenza?
- Nella deduzione, si è detto, uso le leggi del CdP,
- nel passare dal campione alla popolazione ci sono diversi metodi, basati sul CdP.
- Non esiste un unico paradigma dell'inferenza statistica: diversi metodi possono anche, a partire dagli stessi dati e modelli, portare a risposte differenti.
- Una lista, non esaustiva, di metodi per la ricerca di uno stimatore o di un test è
  - metodo basato sulla verosimiglianza;
  - metodo dei momenti;
  - metodo dei minimi quadrati;
  - metodi robusti;
  - metodo bayesiano;
  - bootstrap;
  - ...

Schema delle fasi e modi dell'inferenza statistica