Lezione 6
Metodo della Massima Verosimiglianza

Stimatori di Massima Verosimiglianza

- Un modo molto importante di cercare uno stimatore "buono" è quello di utilizzare il Principio della **Massima Verosimiglianza** o **Maximun Likelihood** (ML):
  Si abbia un campione di n misure $x_1, x_2, \ldots, x_n$ i.i.d. . La LF è una funzione del parametro θ (o dei parametri θ) da stimare: L = L(θ). Lo stimatore di maximun likelihood (**MLE**) $\tilde{\theta}$ è definito come il valore del parametro θ che massimizza la LF per tutti i possibili valori di θ:

$$L(x_1, \cdots, x_n; \tilde{\theta}) \geq L(x_1, \cdots, x_n; \theta)$$

- Se LF è 2 volte differenziabile rispetto a θ, allora bisogna cercare gli zeri della derivata prima rispetto a θ della LF e controllare che la derivata seconda in questo punto sia negativa. Se ci sono più massimi locali si prende il maggiore.

- È molto più comodo invece che utilizzare la LF (dove appare una produttoria) utilizzare il logaritmo della LF (log L) (dove la produttoria diventa una sommatoria). Se la LF è una funzione monotona, i massimi di L sono gli stessi di quelli della log L

Stimatori di Massima Verosimiglianza

- Gli zeri della derivata di log L rispetto a ϑ talvolta si riescono a calcolare per via analitica ma molto più spesso sono determinati per via numerica.

Esempio-1:

- Sia $(x_1, x_2, .. , x_n)$ un campione di dati che segua una distribuzione gaussiana con media μ e varianza σ^2 **parametri non noti** e da determinare.
  Calcoliamo la LF per queste n misure: $L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$

- Passando alla log L :

$$\log L(x_1, \cdots, x_n; \mu, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

e differenziando rispetto a μ e a σ^2, si ottiene:

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu) = 0$$

Stimatori di Massima Verosimiglianza

- Quindi i MLE della media e varianza sono: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$

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

- La media data da questo stimatore è quella aritmetica (non distorta)

- La varianza invece è distorta (ma asintoticamente non distorta). Possiamo comunque correggere la distorsione utilizzando la varianza del campione

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

- Si osservi che $\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{n}{\sigma^2}$

MLE raggiunge il limite di Cramer-Rao

Stimatori di Massima Verosimiglianza

Esempio-2

- Un campione di n misure (x_1, x_2, …, x_n) di una variabile casuale X distribuita seconda una poissoniana:

$$f(x;\theta) = \frac{\theta^x}{x!}e^{-\theta}$$

con θ>0, parametro ignoto da stimare e x = 0, 1, 2 ,…..

- La LF è data da

$$L(\theta) = L(x_1, x_2, .., x_n; \theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

$$logL(\theta) = -n\theta + \left(\sum_{i=1}^{n} x_i\right) log\theta - log\prod_{i=1}^{n} x_i!$$

- Derivando rispetto a θ si ha: $\frac{\partial logL(\theta)}{\partial\theta} = -n + \frac{1}{\theta}\sum_{i=1}^{n} x_i = 0$

da cui $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (la derivata seconda in questo punto è negativa)

- Anche in questo caso Il MLE è non distorto e raggiunge il limite di Cramer-Rao

Stimatori di Massima Verosimiglianza

- Vi possono essere diversi stimatori consistenti e non distorti dello stesso parametro. Ad esempio la mediana di un campione è il valore del parametro che divide il campione in due parti uguali.

- Se la distribuzione delle misure è simmetrica allora la mediana è uno stimatore consistente e non distorto della media del campione.

- Se la distribuzione è gaussiana (media μ e varianza σ^2) allora la varianza sulla media aritmetica $V[\bar{x}]$ e quella sulla mediana $V[\tilde{x}]$ sono:

$$V[\bar{x}_n] = \frac{\sigma^2}{n} \qquad V[\tilde{x}_n] = \frac{\sigma^2}{n} \cdot \frac{\pi}{2}$$

- Lo stimatore mediana ha una varianza maggiore dello stimatore della media però è più **robusto** (cioè è meno influenzato dalle misure nella coda !!)

Invarianza per Trasformazione Funzionale

- MLE è invariante per trasformazione funzionale:
  Se T è MLE di θ e se u(θ) è una funzione di θ, allora u(T) è MLE di u(θ)

- Questa proprietà, molto importante, non è valida per tutti gli stimatori. Ad esempio se T è uno stimatore non distoro di θ, non segue affatto che T^2 sia uno stimatore non distorto di θ^2. Se lo stimatore è il ML, questo e vero!

- Quando possibile noi adottiamo il MLE in quanto:
  1- Se esiste uno stimatore che raggiunge la minima varianza, questo lo si ritrova utilizzando il ML. Questa è la ragione fondamentale per l'utilizzo di questa stimatore.
  2- Sotto condizioni abbastanza generali esso è consistente
  3- Al crescere della dimensione del campione, il MLE è non distorto e raggiunge il limite di minima varianza (è cioè anche efficiente)
  4- Al crescere delle dimensioni del campione , il MLE segue una distribuzione gaussiana.

Varianza negli Stimatori di ML

- Dato un campione di dati e stimato un parametro, alla stima del parametro è associata una incertezza.

- Se rifacessi un altro campione di dati (rifacendo l'esperimento) e stimassi di nuovo il parametro, otterrei una diversa stima dello stesso parametro. Rifacendo N volte l'esperimento, avrei N stime del parametro, stime distribuite secondo una certa distribuzione (che tende ad essere gaussiana per N molto grande)

- Vogliamo determinare la varianza (e deviazione standard) di questa distribuzione.

- Esistono diversi metodi per calcolare la varianza

Metodo Analitico

- In taluni tipi di distribuzione è possibile calcolare la varianza per via analitica. Consideriamo per esempio la distribuzione esponenziale
- Stima ML della vita media : $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$
- Per la varianza si ha: $V[\hat{\tau}] = E[\hat{\tau}^2] - (E[\hat{\tau}])^2 = \frac{\tau^2}{n}$
- In questa formula τ è il valore vero incognito. Grazie all'invarianza sotto trasformazione (di MLE) noi al valore vero possiamo sostituire il valore stimato: $V[\hat{\tau}] = \frac{\hat{\tau}^2}{n}$
- La deviazione standard sul valore stimato del parametro è:

$$\hat{\sigma}_{\hat{\tau}} = \frac{\hat{\tau}}{\sqrt{n}}$$

Limite Inferiore di Cramer-Rao

- Assumendo zero bias ed efficienza, la varianza può essere calcolata a partire dal limite inferiore di Cramer-Rao.

- Le derivate seconde possono essere calcolate dai dati sperimentali e nel punto stimato del parametro. Con queste derivate di ottiene l'informazione di Fisher e quindi:

$$V[\hat{\theta}] = -\left(\frac{1}{\frac{\partial^2 \log L}{\partial \theta^2}}\right)\Bigg|_{\theta=\theta_0}$$

- È il metodo usato quando la LF è massimizzata numericamente (usando per esempio Minuit). Quanto detto si generalizza al caso di n parametri.

Metodo Monte Carlo

- Una volta stimato il parametro dalle misure sperimentali, si potrebbe simulare l'esperimento N volte e stimare da queste simulazioni N valori stimati.

- Gli esperimenti si possono simulare o con simulazioni Monte Carlo complete (che quindi tengono conto delle varie correlazioni tra le variabili). Noi in BaBar (in gergo) li chiamiamo "toy experiments")

- Oppure a partire dalle p.d.f. Questo modo è molto più veloce e semplice del precedente ma non tiene conto per esempio delle eventuali correlazioni tra le variabili. Noi sempre in gergo li chiamiamo "pure toy experiments"

- Dalla distribuzione di questi valori si può dedurre la varianza sul parametro stimato

Metodo Grafico

- Consideriamo prima il caso di un singolo parametro e poi di due parametri.

- Sviluppo il logL(θ) in serie di Taylor attorno al valor stimato $\hat{\theta}$ dal ML:

$$\log L(\theta) = \log L(\hat{\theta}) + \left[\frac{\partial \log L}{\partial \theta}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2!}\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \cdots$$

- Il primo termine dello sviluppo è log L_{max} , il secondo termine è nullo ed il terzo può essere riscritto mediante il limite inferiore di Cramer-Rao, ottenendo:

$$\log L(\theta) = \log L_{max} - \frac{(\theta - \hat{\theta})^2}{2\hat{\sigma^2}_{\hat{\theta}}}$$

- e ponendo $\theta - \hat{\theta} = \pm\hat{\sigma}_{\hat{\theta}}$

  si ottiene $\log L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) = \log L_{max} - \frac{1}{2}$

Metodo Grafico

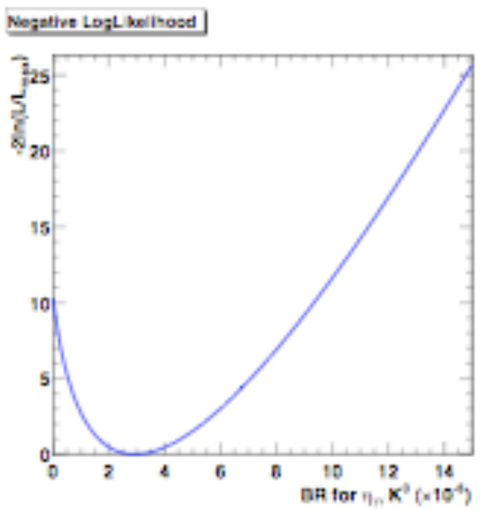
- La relazione appena vista permette di determinare graficamente il valore della deviazione standard. La funzione log L(ϑ), quando cresce la dimensione n del campione di dati, tende a diventare una parabola (in quanto la LF tende ad una gaussiana).

- Invece di massimizzare log L(ϑ) si preferisce minimizzare –log L(ϑ) (log likelihood negativa).

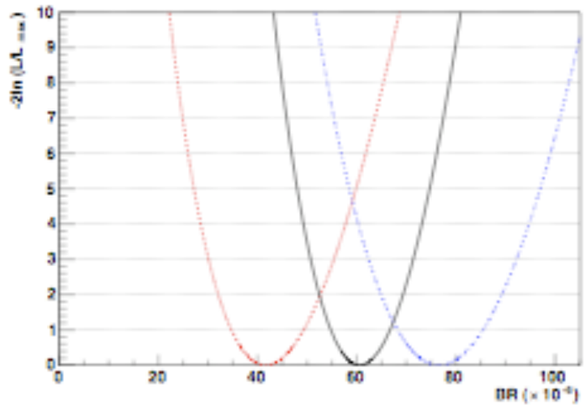
- Anzi ancora meglio si può minimizzare -2logL(ϑ); in questo modo diminuendo di 1 il valore di log L, si ottiene un intervallo di ± σ attorno al valore stimato.

- Si può fare ancora meglio riportando in grafico -2 log(L/Lmax).

Metodo Grafico: esempi



Campione di dati a statistica limitata.
La logL non è una parabola (la LF non è gaussiana). Qui se vi sollevate di 1, avete un intervallo di una σ attorno al valore stimato.
NOTATE: l'intervallo è asimmetrico



Qui il campione di dati ha statistica abbastanza elevata. La logL è parabolica e la LF è gaussiana. Qui l'intervallo ad 1 σ attorno al valore stimato è simmetrico
Le curve rossa e blu corrispondono a due diversi sottodecadimenti dello stesso decadimento del mesone B. La curva nera è la somma delle due logL.

Metodo Grafico (2 Dimensioni)

- Quando i parametri da stimare sono due, la LF $L(\theta_1, \theta_2)$ è una superficie tridimensionale. Per visualizzarla possiamo considerare il luogo dei punti nei quali la logL assume valori costanti (linee di livello).

- Forma della $\log L(\theta_1, \theta_2)$ per grandi campionamenti:

$$\log L(\theta_1, \theta_2) = logL_{max} - \frac{1}{2(1-\rho^2)} \left[\left(\frac{\theta_1 - \hat{\theta}_1}{\sigma_{\hat{\theta}_1}} \right)^2 + \left(\frac{\theta_2 - \hat{\theta}_2}{\sigma_{\hat{\theta}_2}} \right)^2 - 2\rho \left(\frac{\theta_1 - \hat{\theta}_1}{\sigma_{\hat{\theta}_1}} \right) \left(\frac{\theta_2 - \hat{\theta}_2}{\sigma_{\hat{\theta}_2}} \right) \right]$$

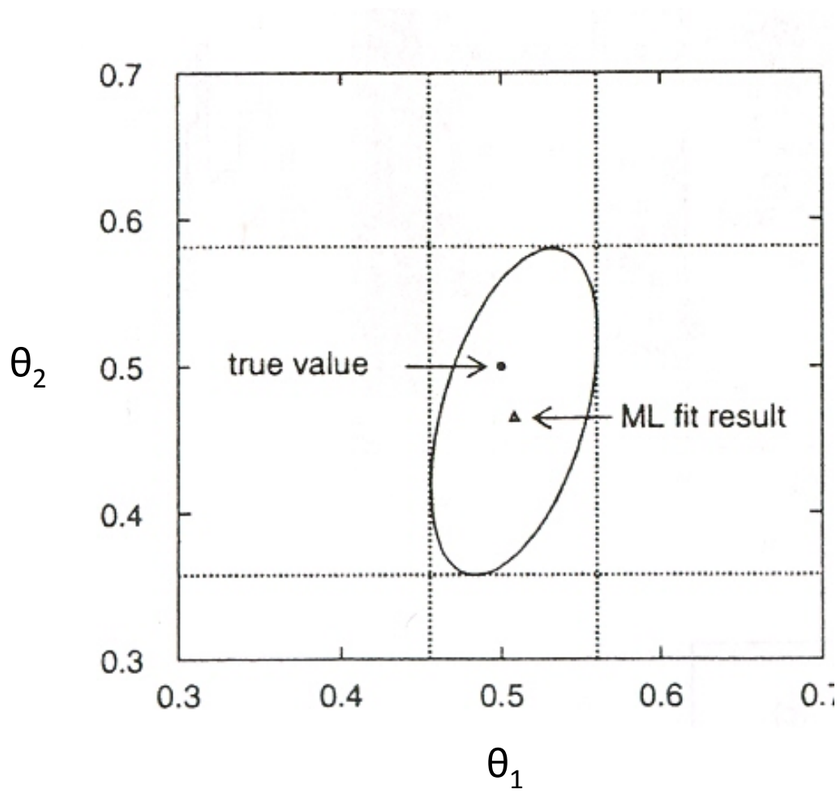
con ρ correlazione tra i valori stimati dei due parametri.

- Il profilo della logL corrispondente a logL = logLmax – 0.5 è dato da:

$$\frac{1}{1-\rho^2} \left[\left(\frac{\theta_1 - \hat{\theta}_1}{\sigma_{\hat{\theta}_1}} \right)^2 + \left(\frac{\theta_2 - \hat{\theta}_2}{\sigma_{\hat{\theta}_2}} \right)^2 - 2\rho \left(\frac{\theta_1 - \hat{\theta}_1}{\sigma_{\hat{\theta}_1}} \right) \left(\frac{\theta_2 - \hat{\theta}_2}{\sigma_{\hat{\theta}_2}} \right) \right] = 1$$

Metodo Grafico (2 Dimensioni)

- La formula vista è l'equazione di una ellisse centrata attorno ai valori stimati. È detta **ellisse di covarianza**



Metodo Grafico (2 Dimensioni)

- L'asse principale dell' ellisse di covarianza forma un angolo α con l'asse θ_1 dato da

$$\tan 2\alpha = \frac{2\rho\sigma_{\hat{\theta}_1}\sigma_{\hat{\theta}_2}}{\sigma^2_{\hat{\theta}_1} - \sigma^2_{\hat{\theta}_2}}$$

- Le tangenti all'ellisse parallele agli assi θ_1 e θ_2 individuano su questi assi intervalli di una σ attorno ai valori stimati.

- Si possono considerare curve di livelli corrispondenti a 2σ, 3σ, ecc.

- La logL ha forma paraboloidale

Maximum Likelihood Estesa

- Noi abbiamo applicato la ML ad un campione di n dati. Molto spesso il numero di eventi n è esso stesso una variabile casuale di tipo poissoniano per esempio con valore medio v.

- Se prendessimo dati per lo stesso periodo di tempo, osserveremmo un numero eventi n' diverso da n (entrambi estratti da una distribuzione poissoniana di media v). Nel nostro stimatore ML vogliamo tener conto che n è una variabile poissoniana (considerando anche la probabilità di osservare n eventi secondo una distribuzione di Poisson) . Quindi scriviamo la LF cosi:

$$L(\nu,\theta) = \frac{\nu^n}{n!}e^{-\nu}\prod_{i=1}^{n} f(x_i;\theta) = \frac{e^{-\nu}}{n!}\prod_{i=1}^{n} \nu f(x_i;\theta)$$

- Questa è detta Maximum Lilelihood Estesa (EML)

Maximum Likelihood Estesa

- Due casi interessanti: ν= ν(θ) oppure ν variabile indipendente da θ

- Consideriamo il primo caso e calcoliamo il logaritmo della LF:

$$logL(\theta) = nlog\nu(\theta) - \nu(\theta) + \sum_{i=1}^{n} logf(x_i;\theta) = -\nu(\theta) + \sum_{i=1}^{n} log(\nu(\theta) f(x_i;\theta))$$

- I termini che non contengono i parametri sono stati eliminati. In questo caso l'inclusione della variabile ν in generale porta ad una diminuzione della varianza dello stimatore

- Consideriamo ora il caso che ν e θ siano indipendenti. Calcolando il logaritmo della LF rispetto a ν e uguagliando a zero, si trova che lo stimatore di ν dà proprio n come valore stimato. Analogamente derivando rispetto a θ si trova un valore stimato identico a quello trovato con la usuale ML.

- Vediamo ora un esempio in cui anche in questa seconda ipotesi l'uso della EML risulta utile

Maximum Likelihood Estesa

- Supponiamo che la p.d.f. di una variabile X sia somma di m contributi

$$f(x;\theta) = \sum_{i=1}^{m} \theta_i f_i(x)$$

con θ_i relativo contributo della componente i-sima. θ_i , parametri da stimare nel fit , sono legati dalla condizione che la loro somma deve essere uguale ad 1 : $\theta_1 + \theta_2 + \ldots\ldots + \theta_m = 1$

- Potrei stimare m-1 parametri e ottenere l'm-esimo parametro da questa condizione. Questo parametro è trattato diversamente dagli altri

- Prendiamo il logaritmo della EML: $logL(\nu,\theta) = -\nu + \sum_{i=1}^{n} log\left(\sum_{j=1}^{m} \nu\theta_j f_j(x_i)\right)$

Ponendo $\mu_i = \theta_i \nu$ abbiamo:

$$logL(\mu) = -\sum_{j=1}^{m} \mu_j + \sum_{i=1}^{n} log\left(\sum_{j=1}^{m} \mu_j f_j(x_i)\right)$$

- Tutti i parametri μ_i sono cosi trattati allo stesso modo

ML con Dati Istogrammati

- Istogrammare i dati è un comodo modo per visualizzare i dati ma si perde informazione (il bin ha dimensione finita)

- Sia n_{tot} il numero totale di eventi del nostro campione di misure di una variabile casuale X, distribuita secondo una p.d.f. $f(x; \theta)$ con $\theta = \theta_1, .. \theta_m$ parametri da stimare.

- Sia n_i il numero di eventi nel bin i-esimo. Allora il numero di eventi aspettati nel bin i è dato da :
  con x_i^{min} e x_i^{max} limiti del bin i.

$$\nu_i(\theta) = n_{tot} \int_{x_i^{min}}^{x_i^{max}} f(x;\theta)dx$$

- Sia N il numero di bin. L'istogramma può essere visto come la misura di una variabile casuale di N dimensioni per il quale la p.d.f. congiunta è una distribuzione multinomiale:

$$f_{congiunta}(\mathbf{n};\nu) = \frac{n_{tot}!}{n_1! \cdots n_N!} \left(\frac{\nu_1}{n_{tot}}\right)^{n_1} \cdots \left(\frac{\nu_N}{n_{tot}}\right)^{n_N}$$

ML con Dati Istogrammati

- La probabilità di essere nel bin i-esimo è data data ν_i/n_{tot}
- Passando ai logaritmi si ha: $logL(\theta) = \sum_{i=1}^{N} n_i log\nu_i(\theta)$
- I parametri da stimare si ottengono massimizzando logL(θ). Questo tipo di ML si dice istogrammato (quello in cui tutti gli eventi sono presi singolarmente si dice ML non istogrammato).
- Quando la larghezza del bin tende a zero il ML istrogrammato tende a quello non istogrammato
- Una grande quantità di misure permette di fare ML istogrammati (molto più semplici) che danno gli stessi risultati dei ML non istogrammati

ML Estesa con Dati Istogrammati

- n_{tot} come variabile poissoniana con media ν_{tot}

$$f_{congiunta}(\mathbf{n};\nu) = \frac{\nu_{tot}^{n_{tot}} e^{-\nu_{tot}}}{n_{tot}!} \frac{n_{tot}!}{n_1!,\cdots,n_N!} \left(\frac{\nu_1}{\nu_{tot}}\right)^{n_1},\cdots,\left(\frac{\nu_N}{\nu_{tot}}\right)^{n_N}$$

con $\nu_{tot} = \Sigma\nu_i$ e $n_{tot} = \Sigma\, n_i$. Utilizzando queste relazioni segue che :

$$f_{congiunta}(\mathbf{n};\nu) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$$

con

$$\nu_i(\nu_{tot},\theta) = \nu_{tot} \int_{x_i^{min}}^{x_i^{max}} f(x;\theta)dx$$

- Passando ai logaritmi si ha $\log L(\nu_{tot},\theta) = -\nu_{tot} + \sum_{i=1}^{N} n_i \log \nu_i(\nu_{tot},\theta)$

- Anche qui si ha una varianza minore se tra ν_{tot} e θ esiste una relazione funzionale.

Bontà del Fit

- Una volta stimato un parametro vogliamo vedere come controllare la qualità del risultato ottenuto, cioè quanto è stato buono il fit che ha dato la stima del parametro.

- Per stimare la bontà del fit bisogna introdurre una statistica di test che ci permetta di fare questa stima. Si determina la p.d.f. di questa statistica

- Noto il valore della statistica di test per il nostro fit, si definisce valore-P **(P-value)** la probabilità di avere un valore della statistica di test che corrisponde ad un risultato egualmente compatibile o meno compatibile di quanto effettivamente osservato nel fit sui dati.

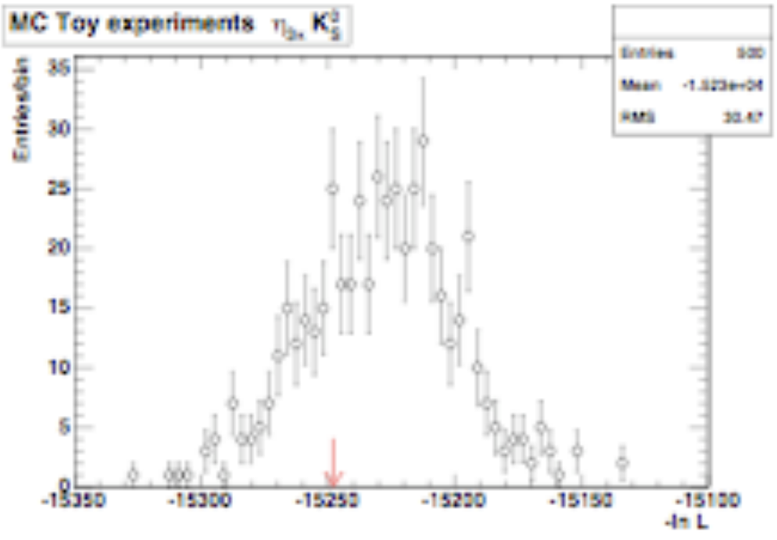
- Il P-value è detto anche livello di significanza o livello di confidenza per il test di bontà del fit

- Vediamo ora come fare un test di bontà del fit col MLE

Bontà del Fit nel ML con $\log L_{max}$

- Uso di logLmax come Statistica di test. Faccio il fit sui dati ed ottengo una stima dei parametri con un certo valore di $-\log L_{max}$ per esempio uguale a -15246. Con tecnica Monte Carlo simulo il mio esperimento un certo numero di volte (per esempio 500 volte). In ogni esperimento simulato faccio il fit per stimare i parametri ed ottengo un valore di $-\log L_{max}$.

- La freccia rossa punta al valore trovato nel fit sui dati. Una volta normalizzata ad 1 questa curva, l'area da questo punto all'infinito fornisce il P-value



Bontà del Fit nel ML con $\log L_{max}$

- Questa tecnica veniva usata sino a qualche anno fa, cioè fino a quando a CDF non hanno scoperto che non c'è alcuna relazione tra bontà del fit e valore della $\log L_{max}$.

- Comunque questo test non è del tutto inutile. Spesso nelle fasi iniziali di una analisi le p.d.f. delle variabili non sono ben sagomate (e talvolta sono sbagliate).

- In questi casi la distribuzione di $-\log L_{max}$ viene del tutto sfasata rispetto alla posizione di $-\log L_{max}$ sui dati. Quindi si corre ai ripari cercando la causa di questo sfasamento.

- Può risultare un utile sanity check !

Bontà del Fit nel MLE con LF

- La statistica di test di bontà del fit può essere basata sulla LF. Date le n_{tot} misure di una variabile casuale X, si istogrammano queste misure e siano $(n_1, n_2, \ldots, n_N)$ in numero di eventi negli N bin. Avendo già stimato i parametri col ML , li utilizzo per determinare i numeri medi di eventi stimati nei vari bin $(\nu_1,\nu_2, .., \nu_N)$

- Questo si può sempre fare anche se il fit di ML è stato non istogrammato

- Uso il numero di eventi dei dati nei bin ed il numero medio di eventi stimati nei bin per costruire una statistica di test di bontà col il rapporto λ :

$$\lambda = \frac{f_{congiunta}(\mathbf{n};\nu)}{f_{congiunta}(\mathbf{n};\mathbf{n})} = \frac{L(\mathbf{n}|\nu)}{L(\mathbf{n}|\mathbf{n})}$$

- Se i dati sono distribuiti secondo una distribuzione poissoniana, allora si ha:

$$\lambda_P = e^{n_{tot}-\nu_{tot}} \prod_{i=1}^{N} \left(\frac{\nu_i}{n_i}\right)^{n_i}$$

Bontà del Fit nel MLE con LF

- Siano m i parametri stimati e supponiamo che la dimensione del campione sia grande. Vedremo tra qualche lezione che in questo caso la variabile $-2\log\lambda_P$ segue una distribuzione del χ^2 con un numero di gradi di libertà pari a N – m (dof = N - m)

$$\chi^2_P = -2\log\lambda_P = 2\sum_{i=1}^{N}\left(n_i \log\frac{n_i}{\hat{\nu}_i} + \hat{\nu}_i - n_i\right)$$

- Se i dati sono distribuiti in modo multinomiale allora si ha:

$$\lambda_M = \prod_{i=1}^{N}\binom{\nu_i}{n_i}^{n_i}$$

e

$$\chi^2_M = -2\log\lambda_M = 2\sum_{i=1}^{N}\left(n_i \log\frac{n_i}{\hat{\nu}_i}\right)$$

- Al limite di grandi campioni anche questa segue una distribuizone del χ^2 ma con N-m-1 gradi di libertà (dof = N-m-1)

Bontà del Fit nel MLE con LF

- Il valore-P (o livello di significanza osservato) è dato da : $P = \int_{\chi^2}^{\infty} f(z; n_d) dz$
  dove f(z; n_d) è la p.d.f. del χ^2 con n_d gradi di libertà

- L'integrale della distribuzione del χ^2 è estesa dal valore di χ^2 osservato sino all'infinito. Valori del χ^2 superiori a quello osservato corrispondono a risultati meno compatibili di quello effettivamente osservato

- Si noti che nella definizione della statistica di test λ il termine al denominatore non dipende dai parametri e potrebbe essere tolto. Si può usare come test di bontà del fit diretamente la likelihood (normalizzata ad 1).

- Per il calcolo di questi P-value si può usare uno dei tanti calcolatori statistici disponibili in rete come ad esempio

  http://www.tutor-homework.com/statistics_tables/statistics_tables.html

  oppure http://socr.ucla.edu/htmls/SOCR_Distributions.html

Bontà del Fit col test di Pearson

- Come nel caso precedente, consideriamo la distribuzione degli n_{tot} eventi osservati in N bin. Nel bin i-esimo abbiano abbiamo n_i eventi osservati e ν_i eventi medi aspettati (utilizzando i valori dei parametri ottenuti dal ML fit). Trattiamo il numero totale di eventi n_{tot} come variabile poissoniana (e quindi variabile)

- Per il test di bontà del fit possiamo usare la statistica del χ^2 (detta di Pearson) :

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \nu_i)^2}{\nu_i}$$

- Se in ogni bin c'è un numero sufficiente di misure (tipicamente ≥ 5) e se in ogni bin il numero di eventi segue una distribuzione di Poisson (ν_i è il valore medio della variabile poissoniana n_i nel bin i-esimo) , allora la statistica di Pearson segue la distribuzione del χ^2 con N-m dof

Bontà del Fit col test di Pearson

- Questa proprietà è sempre vera, indipendentemente dalla forma della distribuzione della variabile X. Per questo motivo il test del χ^2 di Pearson è detto "**distribution free**"

- Se n_{tot} è preso come una costante, allora si ha

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - p_i \cdot n_{tot})^2}{p_i \cdot n_{tot}}$$

con $p_i = v_i / n_{tot}$

- Questa distribuzione segue quella del χ^2 con N-m-1 dof

- Se ci sono bin con meno di 5 eventi, allora le statistiche di Pearson non seguono la distribuzione del χ^2. Questo test non è adeguato per le basse statistiche

- Con basse statistiche si potrebbero usare metodi di simulazione MC ed utilizzare gli esperimenti simulati a maggiore statistica per il test

Combinazione di Più Esperimenti

- Supponiamo che due esperimenti diversi stimino lo stesso parametro, utilizzando eventualmente due variabili diverse X e Y.

- Spesso all'interno della stessa collaborazione gruppi diversi, o anche lo stesso gruppo, misurano lo stesso parametro usando variabili diverse (per esempio si misura il tasso di decadimento di una particella in un certo stato finale utilizzando sottodecadimenti diversi).

- La likelihood totale dei due esperimenti è data da :

$$L(\theta) = \prod_{i=1}^{n} f_1(x_i;\theta) \prod_{i=1}^{m} f_2(y_i;\theta) = L_1(\theta) \cdot L_2(\theta)$$

con $f_1(x;\theta)$, $f_2(y;\theta)$ p.d.f. delle due variabili X e Y e n,m numero di misure nei due esperimenti.

- Se sono note le LF dei due esperimenti possiamo ottenere la LF totale e stimare il parametro θ con una precisione migliore combinando le likelihood

- Questo è il modo più preciso per combinare misure diverse dello stesso parametro

Combinazione di Più Esperimenti

- Questo della likelihood è il modo usuale con cui si combinano diverse misure dello stesso parametro all'interno di una collaborazione

- Sfortunatamente (quasi mai) vengono pubblicate le likelihood

- In pratica quello che si pubblica è il valore del parametro con la sua incertezza che ora supponiamo includa incertezze statistiche e sistematiche. Si usano medie pesate delle misure per combinarle. Per due misure si ha una media:

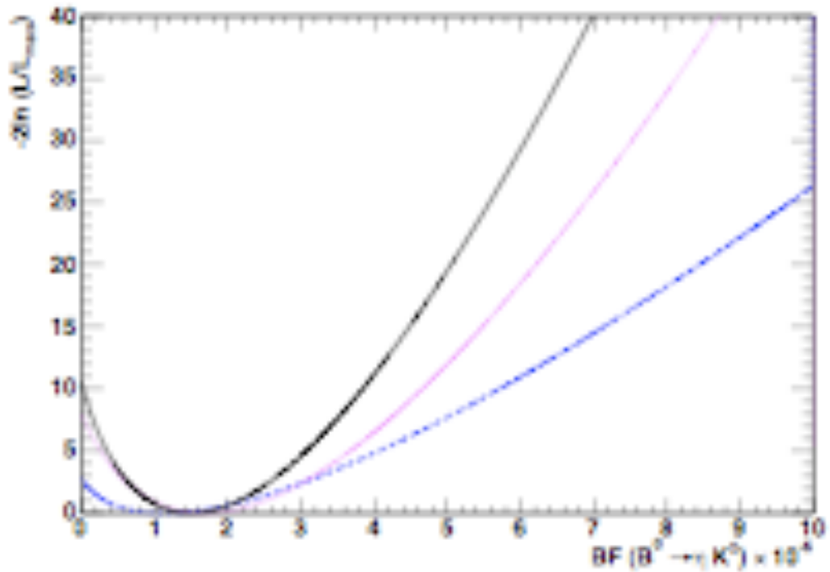
$$\hat{\theta} = \frac{\frac{\hat{\theta}_x}{\hat{\sigma}^2_{\hat{\theta}_x}} + \frac{\hat{\theta}_y}{\hat{\sigma}^2_{\hat{\theta}_y}}}{\frac{1}{\hat{\sigma}^2_{\hat{\theta}_x}} + \frac{1}{\hat{\sigma}^2_{\hat{\theta}_y}}}$$

- Mentre la varianza dello stimatore combinato è: $\hat{V}[\hat{\theta}] = \frac{1}{\frac{1}{\hat{\sigma}^2_{\hat{\theta}_x}} + \frac{1}{\hat{\sigma}^2_{\hat{\theta}_y}}}$

- Particle Data Group (PDG) : http://pdg.lbl.gov

  Heavy Flavor Average Group (HFAG) : http://www.slac.stanford.edu/xorg/hfag

Combinazione di Più Esperimenti



Combinazione di due diverse misure (fatte entrambe da BaBar per determinare il rapporto di decadimento del mesone B^0 -> η K^0 . Curva rosa -2logL per la misura fatta attraverso il sottodecadimento con η -> γγ ; curva blu con η -> 3π. Curva nera : combinazione delle due likelihood (noi come mostrato in figura sommiamo -2 log(L/Lmax). Le likelihood devono includere sia le incertezze statistiche sia quelle sistematiche.

Stimatori Bayesiani

- Una volta fatto un insieme di n misure x_1, x_2, .. , x_n la probabilità bayesiana che un parametro θ assuma un certo valore dipende dalla conoscenza a priori (prior) che abbiamo di quel parametro prima delle misure sperimentali e dalla informazione apportata da queste misure tramite la likelihood.

- Per esempio se devo stimare dalle misure il coseno di un angolo, nel fare il fit la stima bayesiana parte dalla conoscenza a priori cioè che $|\cos\theta| \leq 1$. Questa condizione non viene imposta dalla statistica frequentista dove si estrae il cosθ dalle misure e (a causa di fluttuazioni statistiche) il cosθ può risultare al di fuori dell'intervallo [-1, +1].

- La conoscenza iniziale è contenuta nella distribuzione della prior π(θ)

- La distribuzione bayesiana finale π(θ| x1, … , xn) si ottiene utilizzando il teorema di Bayes: $\pi(\theta \mid x_1, \cdots, x_n) = \frac{\pi(\theta) \cdot L(x_1, \cdots, x_n \mid \theta)}{\int_\Theta \pi(\theta) \cdot L(x_1, \cdots, x_n \mid \theta) d\theta}$

Stimatori Bayesiani

- Per un determinato campione di misure al denominatore della formula di Bayes appare una quantità costante che può essere tolta in quanto le distribuzioni finali sono normalizzate ad 1.

$$\pi(\theta \mid x_1, \cdots, x_n) \sim \pi(\theta) \cdot L(x_1, \cdots, x_n \mid \theta)$$

- La distribuzione bayesiana finale coincide con la likelihood se si assume una prior costante

- La stima puntuale bayesiana del parametro θ si ottiene considerando il valore di aspettazione di θ :

$$\hat{\vartheta} = \int_{\Theta} \theta \cdot \pi(\theta \mid x_1, \cdots, x_n) d\theta$$

- Questo è lo stimatore bayesiano. La sua varianza è la varianza della distribuzione finale.

Stimatori Bayesiani

- Se assumiamo una prior costante, la differenza tra lo stimatore ML e quello bayesiano sta nel fatto che il primo sceglie il valore corrispondente al massimo mentre il secondo ne prende il valore medio (tenendo così conto anche della forma della likelihood)

- Esiste anche la possibilità di definire lo stimatore bayesiano come quello che massimizza la distribuzione finale:

$$\pi(\hat{\theta} \mid x_1, \cdots, x_n) \geq \pi(\theta \mid x_1, \cdots, x_n)$$

per tutti i possibili valori di θ.

- Se si prende costante la prior, allora questo stimatore coinciderebbe con lo stimatore di ML. Comunque resta diversa l'interpretazione del risultato nei due casi.

Stimatori Bayesiani: Esempio

- Una variabile casuale X è distribuita uniformemente nell'intervallo (θ, 10.5). Assumendo che la distribuzione iniziale di θ sia data da:

$$\pi(\theta) = \begin{cases} 5 & \text{per } 9.5 < \theta < 9.7 \\ 0 & \text{altrimenti} \end{cases}$$

determinare la distribuzione finale di θ sapendo che una misura di X è 10.

==========

- Si ha :

$$\pi(\theta|x=10) = \frac{\pi(x=10|\theta)\cdot\pi(\theta)}{\int_{9.5}^{9.7}\pi(x=10|\theta)\pi(\theta)d\theta}$$

$$= \frac{\frac{5}{10.5-\theta}}{\int_{9.5}^{9.7}\frac{5}{10.5-\theta}d\theta}$$

$$= -\frac{1}{(\log 0.8)(10.5-\theta)} \cong \frac{4.48}{10.5-\theta} \qquad \text{per } 9.5 < \theta < 9.7$$

Scelta della Distribuzione Iniziale

- La scelta della distribuzione iniziale riflette le conoscenze che ha lo sperimentatore e quindi è soggettiva. Manca quel carattere oggetivo che è alla base della statistica frequentista.

- La scelta della distribuzione iniziale è in alcuni casi ovvia, in altri molto difficile e in altri casi non si sa come assegnarla.

- Teniamo presente comunque che con grandi campioni di dati la distribuzione finale è largamente dominata dalla likelihood e la scelta della distribuzione iniziale è poco importante

- Una prima prescrizione della distribuzione iniziale venne data dallo stesso Bayes con questo Postulato di Bayes o Principio di Indifferenza :

In assenza di ogni tipo di informazione le distribuzioni iniziali devono essere prese uniformi

Scelta della Distribuzione Iniziale

- Questo postulato sembra quasi ovvio e innocuo ma non è cosi e porta a conclusioni sbagliate.

- Se si assume una prior uniforme per θ, questo in generale non è vero per una una funzione di θ. Presa una prior costante per la frequenza ν la prior per la corrispondente lunghezza d'onda λ = c/ν non è uniforme

- L'uso indiscriminato del postulato di Bayes portò discredito alla impostazione bayesiana della statistica.

- Recentemente si è avuta una rinascita della statistica bayesiana (statistica neobayesiana) con nuove scelte della distribuzione iniziale Il più possibile oggettive (Teoria bayesiana oggettiva)

- La distinzione tra le due statistiche resta comunque netta.