

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

Tecniche statistiche per l'analisi dati della fisica del charm ad LHCb

Relatore:
Dott. Angelo Carbone

Presentata da:
Giovanni Rovinelli

Anno Accademico 2017/2018

Abstract

Il documento riporta un'analisi sull'applicazione di metodi statistici nella determinazione di parametri fisici mediante l'adattamento ai dati di una curva teorica ad una distribuzione di massa invariante del mesone D^0 . Questo studio ha lo scopo di determinare quale metodo statistico è più opportuno impiegare con un'alta statistica di dati, tipica dell'esperimento LHCb che ad oggi e nei prossimi anni dovrà analizzare diversi milioni di dati. In questo studio è stato fatto un confronto tra il metodo di maximum likelihood (ML) e il metodo di χ^2 , sia analizzando una distribuzione binnata di massa invariante sia una non binnata. Le distribuzioni di massa invarianti sono state generate con un metodo Monte Carlo e sono state descritte con una funzione Gaussiana per simulare il segnale e con una funzione esponenziale per descrivere il fondo. Sono stati simulati e analizzati per ogni configurazione 50×10^6 eventi con un rapporto segnale su fondo pari a 1. Nel caso binnato, diverse configurazioni di binning sono state analizzate, in particolare utilizzando da 5 a 1000 bin, nell'intervallo di massa invariante 1810/1920 MeV/ C^2 . E' stato anche valutato il tempo di esecuzione dell'algoritmo di minimizzazione nelle diverse configurazioni. In conclusione si è osservato che mediante i metodi di ML e χ^2 nel caso di analisi di distribuzioni binnate i risultati sono uguali per 1000,500,300, 200, 100, 50 bin e risultano essere uguali anche al caso in cui la distribuzione non è binnata. Nel caso di 10 e 5 bin si sono osservate alcune differenze tra i due metodi nella determinazione dei parametri di interesse, ma anche rispetto ai valori generati. Il tipico tempo di esecuzione di un algoritmo di minimizzazione nel caso di una distribuzione binnata è di pochi secondi, mentre nel caso di una distribuzione non binnata è di circa 8.5×10^3 secondi. Pertanto è possibile concludere che una tipica analisi su un campione di dati di 50×10^6 di eventi di mesoni D^0 può essere fatta sia con il metodo di ML sia con il metodo di χ^2 con una distribuzione di massa invariante binnata, per esempio con 100 bin, o comunque con una configurazione non inferiore di 50 bin.

Indice

1	Fisica delle particelle elementari	6
1.1	Modello Standard	6
1.2	L'interazione Debole	7
1.2.1	Fenomenologia	7
1.2.2	Matrice CKM	9
1.3	Simmetrie	10
1.3.1	Trasformazione di parità e di carica	10
1.4	Violazione CP	10
1.4.1	Violazione di CP nei decadimenti del mesone D^0	11
2	Apparato sperimentale: LHCb	13
2.1	L'acceleratore LHC	13
2.2	L'esperimento LHCb	14
2.3	Tracciamento	15
2.3.1	Il magnete	15
2.3.2	VELO	16
2.3.3	Trigger Tracker	16
2.3.4	Stazioni traccianti	17
2.4	Il sistema di identificazione	18
2.4.1	Rivelatori Cherenkov	18
2.4.2	Calorimetri	18
2.4.3	Camere muoniche	20
2.5	Trigger	20
2.5.1	Livello-0	20
2.5.2	High Level Trigger 1	20
2.5.3	High Level Trigger 2	21
3	Metodi statistici per l'analisi dati	22
3.1	P.D.F. e Maximum Likelihood	22
3.2	Un esempio per ML	23
3.3	ML con distribuzione Gaussiana: stimatori di due parametri	23
3.4	Varianza degli stimatori ML: metodo analitico	24
3.5	Metodo di Monte carlo per la varianza degli stimatori	24
3.6	Metodo grafico per la varianza degli stimatori ML	25
3.7	Metodo dei minimi quadrati	26
3.8	Maximum likelihood con dati binnati	26
3.9	La funzione di Maximum Likelihood Estesa	27

4	Risultati per le simulazioni di Monte Carlo	29
4.1	Modello teorico e parametri della P.D.F. nella simulazione	29
4.2	Simulazione dati	29

Introduzione

Uno degli argomenti più importanti della fisica moderna è lo studio del rapporto tra materia e antimateria. L'esistenza di una differenza tra le due è data dal fatto che vi è una proporzione maggiore della materia rispetto all'antimateria tanto che l'universo sembra essere costituito prevalentemente da materia. Nei primi istanti di vita dell'universo, si pensa che la materia e l'antimateria fossero presenti in proporzioni identiche e che successivamente per via di alcuni processi si è arrivati alla disparità tra le due che notiamo oggi. La trasformazione CP è la traduzione dell'operatore che lega materia e l'antimateria nel Modello Standard. La trasformazione CP è la composizione della trasformazione che concerne la coniugazione di carica e quella di parità. Per diversi anni si è pensato che la trasformazione CP fosse una simmetria valida nella natura; ogni sistema, cioè, si evolve nel suo CP-coniugato. Negli ultimi anni, però, si è osservato come, attraverso esperimenti sempre più accurati e con una strumentazione sempre migliore, in realtà la simmetria CP non sia rispettata. Le particelle e le antiparticelle risultano differenti secondo certi punti di vista. In primis, la scoperta della violazione CP è avvenuta studiando il sistema $K^0 - \bar{K}^0$, ad opera di Fitch, Cronin nel 1964 [1]. Ad oggi sono state osservate violazioni di CP anche nei sistemi $B^0 - \bar{B}^0$ e $B_s^0 - \bar{B}_s^0$. Gli scienziati sono, però, alla ricerca di evidenze della violazione CP nei decadimenti di mesoni contenenti il quark charm; questo perchè il Modello Standard prevede tale asimmetria, osservarla significherebbe dare credito a questa teoria. Quest'asimmetria però si rivela più difficile da notare perchè è attesa essere molto piccola. LHCb ha dato tanta importanza a questo ramo della ricerca portando diversi contributi, grazie anche all'elevata professionalità e al grande apparato di rivelazione che sta alla base di tale esperimento e alla precisione che i macchinari forniscono. In particolare ad LHCb rivelando le collisioni pp (protone-protone) è in grado di studiare il decadimento $D^0 \rightarrow K^+ K^-, \pi^+ \pi^-$ che si ritiene uno dei canali più importanti per la violazione CP da parte dei quark charm. E' possibile studiare la violazione CP attraverso la rivelazione e l'analisi della massa invariante del D^0 ; se, infatti, si ha un tasso del decadimento di un mesone diverso da quello del decadimento del suo CP coniugato, allora si verifica una violazione CP. Per poter effettuare l'analisi della massa invariante nel decadimento del D^0 è necessario, però, l'utilizzo di metodi sulla distribuzione di massa. Tali metodi sono tecniche statistiche che permettono, data una distribuzione di massa dovuta ai dati sperimentali, di calcolare il fit di tale distribuzione e determinare dunque parametri che forniscono l'identità della particella come la media e la deviazione standard per la massa. Il documento è dunque strutturato come segue:

1. Nel primo capitolo vi è una descrizione del Modello Standard, in particolare dell'interazione debole. Si introduce la matrice CKM, che formalizza la violazione CP; viene poi studiata in particolare la violazione CP per il quark charm e attraverso l'utilizzo della definizione di ampiezza di decadimento si fornisce un osservabile che valuta l'asimmetria nel decadimento del D^0 .
2. Nel secondo capitolo viene descritto l'apparato sperimentale LHCb installato al Cern di Ginevra; ci si sofferma oltre che sul detector anche sul sistema di identificazione e tracciamento delle particelle prodotte dalle collisioni.
3. Nel terzo capitolo vengono fornite a livello teorico le tecniche statistiche per lo studio del fit della massa invariante. Si descrive la differenza tra il metodo dei minimi quadrati (con l'uso del χ^2) e quello della massima verosimiglianza (ML, con la funzione di Likelihood). Viene, inoltre, esteso il modello ML in modo da poter considerare una situazione più verosimile

4. Nel quarto capitolo, infine, vengono riportati i risultati delle simulazioni fatte per confrontare i metodi citati nel capitolo 3. Si valuta l'applicabilità di tali metodi e quale di questi è più conveniente per alta statistica. Viene, infine, fatto un confronto tra i tempi di esecuzione del fit notando una discrepanza sostanziale tra i risultati.

Capitolo 1

Fisica delle particelle elementari

1.1 Modello Standard

Negli ultimi cinquant'anni sono stati compiuti notevoli progressi nell'identificazione dei costituenti fondamentali della materia e delle forze fondamentali attraverso le quali questi costituenti interagiscono. Il Modello Standard (MS) è il punto più alto della conoscenza nel campo delle particelle e delle loro interazioni elementari. Sono quattro i tipi di interazioni fondamentali esistenti: forte, debole, elettromagnetica e gravitazionale. Le particelle invece, secondo tale teoria si dividono in due grandi categorie: i fermioni e i bosoni. I fermioni sono suddivisi in:

1. quark, particelle che hanno carica di colore e pertanto le uniche ad interagire fortemente. Sono sei e si dividono in tre generazioni, rispettivamente up-down, charm-strange, top-bottom. Hanno tutte spin $1/2$ e possiedono carica frazionaria, pari a $+2/3$ per i quark up-like e a $-1/3$ per i quark down-like;
2. leptoni, particelle di spin $1/2$ divise anch'esse in tre generazioni, dette dell'elettrone, del muone e del tau, che contengono ciascuna una particella di carica -1 e il corrispondente neutrino, che non ha carica ed è soggetto esclusivamente all'interazione debole.

I fermioni inoltre obbediscono alla statistica di Fermi-Dirac ed hanno spin semintero; viceversa i bosoni si comportano seguendo la statistica di Bose-Einstein ed hanno spin intero. I mediatori dell'interazione debole, forte ed elettromagnetica sono bosoni. I bosoni sono pertanto di tre tipi, poichè sono mediatori di tre tipi di forze e sono:

1. il fotone, particella priva di massa e carica ma con spin 1, mediatrice dell'interazione elettromagnetica;
2. i gluoni, mediatori dell'interazione forte aventi le stesse proprietà del fotone con l'aggiunta della carica di colore, che li separa in 8 differenti particelle;
3. i bosoni massivi, W^+ , W^- , Z^0 , mediatori dell'interazione debole di spin 1 e con carica rispettivamente $+1$, -1 e 0 .

In realtà abbiamo un quarto tipo di bosone che non è associato a nessuna interazione fondamentale ma è necessario perchè conferisce la massa alle particelle elementari: il bosone di Higgs. Quest'ultimo è massivo e scalare e non possiede alcun tipo di carica. Ad ogni particella si accompagna una corrispondente antiparticella, avente la stessa massa e vita media ma tutti i restanti numeri quantici invertiti. Le interazioni fondamentali sono descritte tramite le teorie di Gauge. L'idea di fondo di tali teorie è che ogni particella esista in un opportuno N numero di stati di carica ordinati in un vettore (cariche di gauge) e che lungo l'interazione vi sia una transizione da uno stato di carica all'altro mediata evidentemente da una matrice di ordine N . Una rotazione di tali vettori cariche (cioè l'applicazione di una matrice di trasformazione) non deve produrre osservabili; ciò sta ad indicare l'invarianza rispetto alle rotazioni degli stati di carica ovvero

rispetto a trasformazioni unitarie e unimodulari di rango N , indicate con $SU(N)$. L'intera teoria dinamica riconosce fondamentale il requisito di invarianza rispetto a trasformazioni di tipo $SU(N)$ dipendenti dalla posizione, per cui si parla di teorie costruite con la simmetria per trasformazioni locali di gauge $SU(N)$. Dunque il Modello standard si basa sul gruppo di gauge [2].

$$G_{MS} = SU(3)_C \otimes SU(2)_T \otimes U(1)_Y, \quad (1.1)$$

dove $SU(3)$ indica l'invarianza per l'interazione forte, $SU(2) \otimes U(1)$ rappresenta, invece, l'unificazione dell'interazione debole con quella elettromagnetica (elettro-debole). I pedici indicano i numeri quantici conservati: C , la carica di colore, T l'isospin debole e Y l'ipercarica debole. Inoltre bisogna considerare che vi è un accoppiamento tra carica e campo che prende forma in modo algebrico attraverso un accoppiamento del tipo matrice-vettore. Poichè una matrice unitaria e unimodulare di ordine N è sempre rappresentabile come sovrapposizione di $N^2 - 1$ matrici di ordine N linearmente indipendenti, si deve ammettere l'esistenza di $N^2 - 1$ campi di gauge dotati di carica di gauge e quindi soggetti a mutua interazione. Un esempio di quanto detto è rappresentato dall'interazione forte. Considerando la fenomenologia di tale interazione possiamo confermare la validità delle teorie assumendo che vi siano tre stati di carica per ogni quark, dette cariche di colore ed indicate con i nomi: rosso, blu e verde. Questo comporta che debbano esistere 6 tipi di quark in tre diverse cariche con stessa massa e spin, per un totale di 18 quark. I campi associati alle cariche invece saranno $3^2 - 1 = 8$ campi di gauge, o meglio, campi gluonici, portatori di carica di colore. Per le interazioni forti abbiamo una simmetria rispetto trasformazioni di tipo $SU(3)$. L'interazione debole richiede invece 2 stati di carica, che vengono chiamate cariche di isospin debole (questo perchè si comportano come lo spin). Una differenza sostanziale con l'interazione forte sta nel fatto che i due stati di carica, per l'esigenza di descrivere la fenomenologia, devono essere identificati con particelle aventi diverso sapore e diversa massa ovvero con particelle che sono considerate tra loro diverse. I bosoni accoppiati ai 3 campi deboli (essendo 2 le cariche deboli, $N^2 - 1 = 4 - 1 = 3$) sono W^\pm, Z^0 . Le particelle interagiscono debolmente a coppie. Come coppie di cariche di isospin debole consideriamo le generazioni di particelle leptoniche e di quark:

1. quark-quark: up-down, charm-strange, top-bottom;
2. leptone-neutrino: elettrone- ν_e , muone- ν_μ , tauone- ν_τ .

Di seguito si mostra la totalità delle particelle elementari fino ad ora conosciute (non considerando le antiparticelle corrispondenti).

1.2 L'interazione Debole

1.2.1 Fenomenologia

Come le interazioni forti ed elettromagnetiche, quella debole è anch'essa associata ad un bosone elementare che agisce tra quarks e/o leptoni. Si hanno come detto tre particelle mediatrici dell'interazione, i bosoni W^\pm e Z^0 , con massa pari a circa 80.3 e 91.2 GeV rispettivamente. Il raggio d'azione dell'interazione è molto ridotto, dell'ordine di 10^{-18} ; inoltre essa risulta meno intensa delle interazione elettromagnetica e forte, per cui è normalmente schermata da queste ultime, ma diventa l'interazione principale solo nei processi che per le altre interazioni sono vietati. L'interazione debole può avvenire tra leptoni e quark (interazioni semileptoniche), tra soli leptoni (interazioni leptoniche) o tra soli quark (interazioni non leptoniche). I diagrammi di Feynman ci permettono di rappresentare l'interazione in modo molto immediato. Lungo l'asse orizzontale è rappresentato il tempo; quindi ciò che sta a destra sono le particelle che interagiscono, ciò che sta a sinistra, invece, sono le particelle risultanti e in mezzo abbiamo i due vertici dell'interazione legati dallo scambio di un bosone.

Sono due i tipi di processi che possiamo avere:

QUARKS	UP mass 2,3 MeV/c ² charge 2/3 spin 1/2 u	CHARM 1,275 GeV/c ² 2/3 1/2 c	TOP 173,07 GeV/c ² 2/3 1/2 t	GAUGE BOSONS	GLUON 0 0 1 g	HIGGS BOSON 126 GeV/c ² 0 0 H
	DOWN 4,8 MeV/c ² -1/3 1/2 d	STRANGE 95 MeV/c ² -1/3 1/2 s	BOTTOM 4,18 GeV/c ² -1/3 1/2 b		PHOTON 0 0 1 γ	
	LEPTONS				Z BOSON 91,2 GeV/c ² 0 1 Z	
					W BOSON 80,4 GeV/c ² ±1 1 W	
	ELECTRON 0,511 MeV/c ² -1 1/2 e	MUON 105,7 MeV/c ² -1 1/2 μ	TAU 1,777 GeV/c ² -1 1/2 τ			
	ELECTRON NEUTRINO <2,2 eV/c ² 0 1/2 ν_e	MUON NEUTRINO <0,17 MeV/c ² 0 1/2 ν_μ	TAU NEUTRINO <15,5 MeV/c ² 0 1/2 ν_τ			

Figura 1.1: particelle elementari

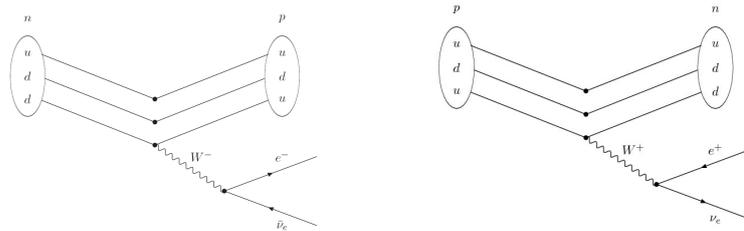


Figura 1.2: Decadimento β , β^- .

1. a corrente carica: Un esempio è il decadimento beta, o beta positivo che scambiano W^- o W^+ ,
2. a corrente neutra: il bosone scambiato è Z^0 , la reazione è: $\nu\mu + N \rightarrow \nu\mu + X$, dove N è un nucleone e X è un adrone permesso dalla legge di conservazione.

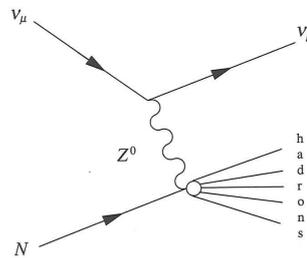


Figura 1.3: diagramma di Feynman per una la reazione neutra di corrente: $\nu\mu + N \rightarrow \nu\mu + X$

1.2.2 Matrice CKM

L'interazione debole dei quark è molto più complicata di quella dei leptoni, la possiamo comprendere meglio con le idee di simmetria leptone-quark e miscelazione dei quark. Se consideriamo il caso di semplici generazioni di quark e di leptoni avremo che le prime due generazioni di quark:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad e \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad (1.2)$$

e le prime due generazioni di leptoni

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \quad e \quad , \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \quad (1.3)$$

hanno l'interazione debole identica. I risultanti vertici quark con W^\pm è mostrato nella seguente figura:

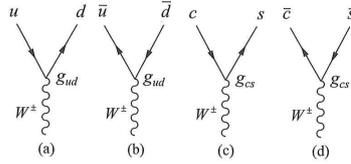


Figura 1.4: Vertici W^\pm -quark ottenuti dalla simmetria quark-leptone e non dal miscelamento

il termine g è il termine d'accoppiamento ed è l'analogo debole della carica e nelle unità appropriate. La simmetria dei quark nella forma semplice implica che i processi fondamentali $d + \bar{u} \rightarrow W^-$ e $s + \bar{c} \rightarrow W^-$ si verificano con lo stesso accoppiamento g_W come il corrispondente processo leptonic. In Figura 1.4 abbiamo $g_{cs} = g_{ud} = g_W$, mentre i processi $s + \bar{u} \rightarrow W^-$ e $d + \bar{c} \rightarrow W^-$ non sono contemplati. Dunque nel Modello Standard è contemplato l'accoppiamento dei quark con l'interazione debole in maniera dipendente dalle generazioni dei quark coinvolti nel processo. Questa idea venne proposta inizialmente da Cabibbo, che la formalizzò per i 3 quark conosciuti all'epoca, ipotizzando che gli stati soggetti ad interazione debole non siano gli autostati di massa u, d e s , ma piuttosto degli stati "ruotati" u', d' ed s' tali che $u' = u$ e

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} \cos(\theta_c) & \sin(\theta_c) \\ -\sin(\theta_c) & \cos(\theta_c) \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix} \quad (1.4)$$

dove θ_c è un parametro detto angolo di Cabibbo, che quantifica la rotazione degli stati rispetto agli autostati di massa, e la matrice di rotazione appena presentata è chiamata matrice di Cabibbo [3]. Studiando processi dovuti ad interazione debole che coinvolgono questi 3 quark si può ottenere una stima dell'angolo di miscelamento, che risulta essere $\theta_c = 12.8^\circ$. Da questo valore appare evidente che i processi che non violano la stranezza ($\Delta S = 0$) sono favoriti rispetto ai processi che la violano ($\Delta S \neq 0$), in quanto i primi saranno proporzionali a $\cos(\theta_c)$, mentre i secondi a $\sin(\theta_c)$.

Questo principio può essere esteso per incorporare i restanti 3 quark, non noti al tempo di Cabibbo. La trattazione attuale è derivata dal lavoro di Kobayashi e Maskawa, i quali operarono la sostituzione delle 3 generazioni di quark soggette ad interazione forte con le 3 generazioni ruotate, autostati dell'interazione debole:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \Rightarrow \begin{pmatrix} u \\ d' \end{pmatrix} \begin{pmatrix} c \\ s' \end{pmatrix} \begin{pmatrix} t \\ b' \end{pmatrix}, \quad (1.5)$$

per cui i quark carichi positivi restano invariati e quelli carichi negativi sono dati da

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \quad (1.6)$$

V_{CKM} è la cosiddetta matrice di Cabibbo–Kobayashi–Maskawa (o CKM). Si tratta di una matrice complessa unitaria che codifica il miscelamento degli autostati di massa dei quark a formare gli autostati dell'interazione debole. La matrice fornisce le probabilità di transizione da un quark q ad un altro quark q' , che sono proporzionali a $|V_{qq'}|^2$. Le transizioni con maggiore probabilità sono quelle tra quark di una stessa generazione (u-d, c-s, t-b), mentre quelle tra quark di generazioni differenti sono sopresse di termini crescenti con l'aumentare della distanza tra le stesse (la catena di decadimenti favorita è $t \rightarrow b \rightarrow c \rightarrow s \rightarrow u$). Gli elementi della matrice CKM possono essere ricavati dallo studio di particolari processi deboli, e i loro valori aggiornati risultano

$$\begin{pmatrix} 0.97434 & 0.22506 & 0.00357 \\ 0.22492 & 0.97351 & 0.0411 \\ 0.00875 & 0.0403 & 0.99915 \end{pmatrix}. \quad (1.7)$$

1.3 Simmetrie

1.3.1 Trasformazione di parità e di carica

In fisica, per parità si intende la proprietà di un fenomeno di ripetersi immutato dopo un'inversione delle coordinate spaziali. Quando avviene tale fenomeno si dice che la parità è conservata, altrimenti è violata. In realtà essa si conserva per l'elettromagnetismo, la forza gravitazionale e l'interazione forte, ma non per le interazioni deboli. Siano $r=(x,y,z)$ le coordinate spaziali di un punto e $|\psi(r, t)\rangle$ uno stato, allora l'inversione è descritta da un operatore \hat{P} tale che:

$\hat{P}|\psi(r, t)\rangle = |\psi(-r, t)\rangle = |\psi(-x, -y, -z, t)\rangle$. Applicando due volte l'operatore si ottiene lo stato di partenza, questo perché $\hat{P}^2 = I$, per cui \hat{P} può avere gli autovalori $P = \pm 1$. Questo significa che uno stato può avere parità definita positiva ($P = 1$) oppure negativa ($P = -1$) [3]. In un qualsiasi processo fisico, il numero quantico di parità è conservato se l'operatore parità commuta con l'hamiltoniana, $[\hat{H}, \hat{P}] = 0$. La trasformazione di carica, invece, anche detta coniugazione di carica consiste nell'invertire tutti i numeri quantici di una particella, a parte lo spin. Con l'inversione di carica una particella diventa un'antiparticella. Anche in questo caso possiamo descrivere la trasformazione con un operatore: \hat{C} . Esso di nuovo agisce su uno stato $|\psi(s, q)\rangle$, di spin s e cariche q in modo che:

$$\hat{C}|\psi(s, q)\rangle = |\psi(s, -q)\rangle.$$

L'autovalore di \hat{C} può assumere soltanto i valori $C = \pm 1$. Questo significa che la particella può essere autostato di \hat{C} solo se coincide con la sua antiparticella (vedi il fotone γ). È doveroso osservare che l'aver carica elettrica nulla non è una condizione sufficiente per essere autostato di \hat{C} [2].

1.4 Violazione CP

Per molto tempo si è pensato che la parità e la C-parità fossero sempre conservate in un qualsiasi processo, finché nel 1957 madame Wu mostrò che nel decadimento β del neutrone, mediato dall'interazione debole, la conservazione della parità viene violata. Si osservò che applicando la coniugazione di carica ad un neutrino, che si sapeva essere sinistrorso [2], lo si trasforma in un antineutrino sinistrorso, che non è mai stato osservato. Ciò mostra che nemmeno la C-parità è conservata dall'interazione debole. Tramite la composizione delle due trasformazioni CP, però, un neutrino sinistrorso viene trasformato in un antineutrino destrorso. Questo portò a pensare che anche se vengono violate la parità e la C-parità, vi sia una conservazione per l'interazione debole di CP. Cronin e Fitch, però, nel 1964, osservarono la violazione della simmetria CP nel decadimento di K_L^0 . Iniziò di lì una ricerca scientifica importante sulla violazione della simmetria CP, che portò alla distinzione di due tipi di violazione:

1. diretta, quando la particella si comporta diversamente dall'antiparticella (diversi tassi di decadimento)
2. indiretta, quando la probabilità di una particella neutra di oscillare in antiparticella è diversa dalla probabilità dell'antiparticella di oscillare in particella.

1.4.1 Violazione di CP nei decadimenti del mesone D^0

Una delle violazioni della simmetria CP, come detto precedentemente è quella per cui si ha una differenza tra la probabilità di transizione da mesoni ad antimisone e la probabilità del processo inverso. Questo tipo di violazione CP è indipendente dal modo di decadimento. Quando si considera anche il decadimento di un mesone, la violazione CP è data dalla diseguglianza di ampiezza di decadimento di un mesone allo stato finale f , $|A_f|$, confrontata con l'ampiezza del decadimento dell'anti-mesone allo stato di violazione CP \bar{f} , $|\bar{A}_{\bar{f}}|$ [4]. Dunque la violazione CP si ha quando $|\bar{A}_{\bar{f}}/A_f| \neq 1$. Questo modo, detto violazione diretta di CP, prevede il parametro

$$a_d \equiv \frac{|A_f|^2 - |\bar{A}_{\bar{f}}|^2}{|A_f|^2 + |\bar{A}_{\bar{f}}|^2}. \quad (1.8)$$

Infine, per decadimenti di mesoni neutri il miscelamento e l'ampiezze di decadimento possono interferire; si ha un terzo tipo di violazione CP. Questa violazione nell'interferenza è presente nel numero complesso definito come

$$\lambda_f = \frac{q\bar{A}_f}{pA_f} = -\eta_{CP} \left| \frac{q}{p} \right| \left| \frac{\bar{A}_f}{A_f} \right| e^{i\phi}. \quad (1.9)$$

L'espressione è valida per un autostato f , con autovalore η_{CP} e ϕ è la relativa fase della violazione CP tra q/p e \bar{A}_f/A_f . La combinazione della violazione CP in interferenza e miscelamento è chiamata anche violazione indiretta di CP. In generale c'è violazione della simmetria CP se $\lambda_f \neq 1$. Diversamente dallo strange e del beauty, la violazione CP non è ancora stata scoperta nei sistemi contenenti il charm. E' possibile scrivere l'ampiezza di decadimento come

$$A_f = C(1 + r_f e^{t(\delta_f + \phi_f)}), \quad (1.10)$$

dove r_f è la grandezza dell'ampiezza e la relativa fase forte e debole sono δ_f e ϕ_f . Corrispondentemente, il decadimento CP-coniugato è dato da

$$\bar{A}_{\bar{f}} = C(1 + r_f e^{t(\delta_f - \phi_f)}), \quad (1.11)$$

in cui la fase debole cambia segno mentre la fase forte no. Usando il tempo di tasso di decadimento

$$\Gamma(M \rightarrow f) = \int_0^\infty \Gamma(M(t) \rightarrow f) dt \propto |A_f|^2, \quad (1.12)$$

si ottiene per l'asimmetria CP

$$A_{CP} \equiv \frac{\Gamma(M \rightarrow f) - \Gamma(\bar{M} \rightarrow \bar{f})}{\Gamma(M \rightarrow f) + \Gamma(\bar{M} \rightarrow \bar{f})} = 2r_f \sin(\delta_f) \sin(\phi_f). \quad (1.13)$$

Questa equazione mostra che la violazione CP può verificarsi solo nel caso che due o più ampiezze contribuiscono al decadimento ($r_f \neq 0$) con diversa fase forte e debole [4]. Le due ampiezze richieste per la violazione CP possono per esempio essere date dal decadimento a tre livelli o quello e dal decadimento a pinguino come si vede in Figura 1.5 . Il diagramma a pinguino contiene una coppia quark/antiquark dello stesso colore. Per il charm le transizioni sopresse sono $c \rightarrow u\bar{d}d$ e $c \rightarrow u\bar{s}s$, che si verificano per esempio nel decadimento $D^0 \rightarrow \pi^- \pi^+$ e $D^0 \rightarrow K^- K^+$. La violazione di CP diretta è l'asimmetria tra il tasso di decadimento e il tasso di decadimento del suo CP coniugato, come definito dalla relazione

$$A_{CP}(D \rightarrow f) \equiv \frac{\Gamma(D \rightarrow f) - \Gamma(\bar{D} \rightarrow \bar{f})}{\Gamma(D \rightarrow f) + \Gamma(\bar{D} \rightarrow \bar{f})}, \quad (1.14)$$

per un decadimento di un mesone D ad uno stato finale f . E' possibile, dunque, osservare la differenza nella simmetria CP dei decadimenti dei mesoni D^0 in due cariche di pioni o di kaoni. In questo caso tutte le asimmetrie si cancellano al primo ordine e la differenza misurata è uguale alla differenza:

$$\Delta A_{CP} \equiv a_{CP}(K^- K^+) - a_{CP}(\pi^- \pi^+) - a_{raw}(\pi^- \pi^+). \quad (1.15)$$

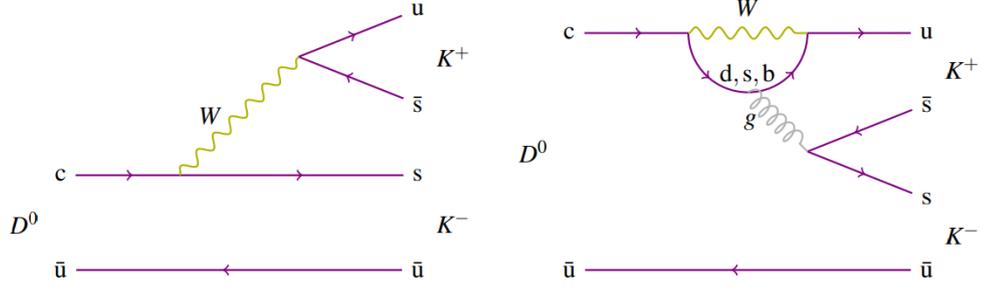


Figura 1.5: a sinistra un diagramma a tre livelli, a destra un diagramma a pinguino per il decadimento $D^0 \rightarrow K^- K^+$

Tra tutti i possibili decadimenti, i decadimenti $D^0 \rightarrow K^- K^+$ e $D^0 \rightarrow \pi^- \pi^+$, sono, dunque, i canali più promettenti per la ricerca di una violazione CP. Sperimentalmente, un importante osservabile per la ricerca della violazione CP diretta è proprio ΔA_{CP} . Le misure del precedente osservabile sono state realizzate da CDF, BaBar, e la collaborazione Belle ([5], [6], [7]), e da LHCb. Gli attuali risultati sperimentali sono stati riassunti in Tabella 1.

Esperimenti	ΔA_{CP}	Referenze
CDF	$(-62 \pm 21 \pm 10) \times 10^{-4}$	[5]
BaBar	$(+24 \pm 62 \pm 26) \times 10^{-4}$	[6]
Belle	$(-87 \pm 41 \pm 6) \times 10^{-4}$	[8]
LHCb ($3.0 fb^{-1}$, <i>muone</i>)	$(14 \pm 16 \pm 8) \times 10^{-4}$	[9]
LHCb ($3.0 fb^{-1}$, <i>pione</i>)	$(-10 \pm 8 \pm 3) \times 10^{-4}$	[10]

Tabella 1. Risultati sperimentali di ΔA_{CP} . Per il riferimento [6], l'asimmetria CP per $D^0 \rightarrow K^- K^+, \pi^+ \pi^-$ è stata calcolata separatamente; così Δ_{CP} è stata calcolata con un'incertezza sommata in quadratura.

Capitolo 2

Apparato sperimentale: LHCb

2.1 L'acceleratore LHC

Il Large Hadron Collider è il più grande acceleratore di particelle del mondo. Situato nei pressi di Ginevra, al confine tra Francia e Svizzera, è costituito da un anello acceleratore di 26.7 km di circonferenza, e ospita 4 grandi esperimenti di fisica delle particelle:

1. ATLAS e CMS: rivelatori general purpose ermetici, pensati per individuare tutti i prodotti delle collisioni e studiare in modo generale le proprietà delle interazioni fondamentali ad alte energie. I loro obiettivi sono la verifica del Modello Standard, la ricerca del bosone di Higgs, recentemente scoperto, e la fisica oltre il Modello Standard;
2. LHCb: esperimento dedicato allo studio della violazione della simmetria CP nei decadimenti di mesoni B e D; specializzato in particolare nel settore della fisica del sapore.
3. ALICE: detector progettato per studiare urti tra ioni pesanti (tipicamente piombo-piombo ma anche protone-piombo) allo scopo di realizzare il cosiddetto plasma di quark e gluoni, uno stato della materia che si pensa essere esistito solo nei primissimi istanti di vita dell'universo.

In Figura 2.1 è possibile vedere l'estensione dell'apparato. LHC è in grado di far collidere due fasci di protoni fino ad un'energia nel centro di massa di 14 TeV, con una luminosità massima di progetto di $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. I fasci sono mantenuti stabili da 1232 magneti superconduttori raffreddati a 1.9°K che esercitano un intenso campo magnetico da 8 T. I protoni viaggiano lungo l'acceleratore in pacchetti (bunches) di 10^{11} particelle che, durante la fase operativa, si scontrano ad una frequenza di 40 MHz, generando circa 20 collisioni per ciascun incontro dei fasci.

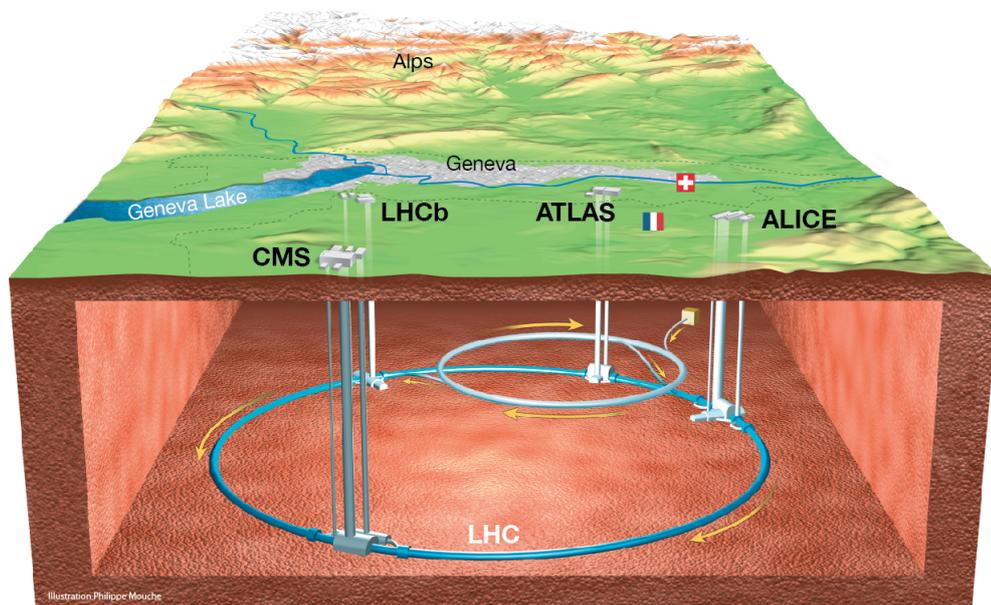


Figura 2.1: LHC e i 4 esperimenti

2.2 L'esperimento LHCb

Il rivelatore LHCb è costituito da uno spettrometro a singolo braccio proiettato in avanti rispetto al punto di interazione. Questo design è stato scelto in base a due proprietà della produzione di quark pesanti:

1. Le coppie di quark $b\bar{b}$ e $c\bar{c}$ vengono prodotte prevalentemente in avanti, ossia parallelamente alla linea dei fasci; un detector con una piccola copertura angolare è dunque in grado di rilevare gran parte delle coppie prodotte.
2. tali particelle, quando prodotte, presentano un notevole boost lorentziano lungo la loro direzione di volo; ciò gli permette di coprire distanze mediamente più lunghe di quelle percorse da particelle di altro tipo. Questa particolarità gioca un ruolo importante nella risoluzione dei tempi di decadimento e rende più facile distinguere le particelle secondarie da quelle prodotte nelle collisioni iniziali.

La geometria del rivelatore è mostrata in Figura 2.2.

Definito un sistema di coordinate destrorso (x, y, z) in cui l'asse z è parallelo alla linea dei fasci e l'asse y è orientato verso l'alto, i sottorivelatori sono posizionati in sequenza lungo la linea dei fasci. L'apparato copre una regione angolare compresa tra 10 e 300 mrad nel piano di deflessione $(x-z)$ e tra 10 e 250 mrad nel piano di non deflessione $(y-z)$. La luminosità istantanea del rivelatore è mantenuta intorno a $2 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, un valore circa 50 volte inferiore alla massima luminosità raggiungibile. Questo perché LHCb è progettato per lavorare in maniera ottimale in condizioni di ridotta molteplicità di interazioni sovrapposte, per meglio determinare la vita media dei mesoni e il loro stato di sapore. Inoltre un minor numero di interazioni diminuisce la velocità con cui le componenti del rivelatore vengono danneggiate a causa della dose di radiazione assorbita. Il numero medio di interazioni per incrocio dei fasci può essere variato regolando le ottiche di questi ultimi nel punto di interazione; in questo modo la luminosità istantanea può essere mantenuta pressoché costante durante la presa dati, nonostante l'intensità dei fasci diminuisca col passare del tempo per colpa delle numerose collisioni. L'esperimento è formato da numerosi

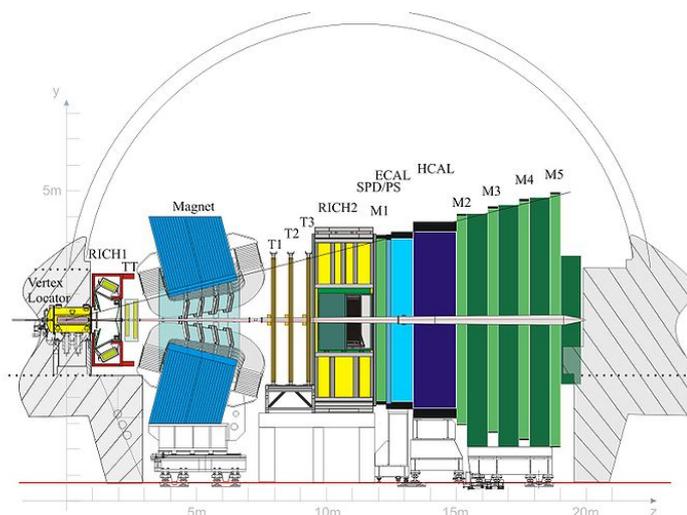


Figura 2.2: Geometria del rivelatore

apparatati in grado di ricavare informazioni sull'identità, la traiettoria, l'impulso e l'energia delle particelle prodotte, distinguendo ognuna di esse dalle altre centinaia di particelle generate nel punto di interazione. I sottorivelatori possono essere raggruppati in due categorie:

1. Il sistema di tracciamento è composto dal localizzatore di vertice (VELO), che circonda la zona d'interazione, e le stazioni traccianti, una davanti (il Trigger Tracker) e tre dietro (T1-T3) il magnete.
2. Il sistema di identificazione delle particelle si compone dei due rivelatori Ring-Imaging Cherenkov, posti prima e dopo il magnete (RICH 1 e RICH 2), dei due calorimetri elettromagnetico ed adronico (ECAL e HCAL) assieme ai due rivelatori ausiliari Scintillating Pad Detector (SPD) e Pre-Shower detector (PS), e delle cinque camere muoniche (M1-M5), di cui la prima è posta davanti ai calorimetri e le restanti 4 in fondo al rivelatore.

Di seguito vengono descritte nel dettaglio le componenti appena citate.

2.3 Tracciamento

Le prestazioni del detector dipendono in larga misura dalla capacità di ricostruire con grande precisione i vertici di interazione, le traiettorie e gli impulsi delle particelle. Il sistema tracciante adempie a questo scopo per mezzo del rivelatore di vertice (VERTex LOCator, VELO), del Trigger Tracker (TT) e delle stazioni traccianti T1, T2 e T3.

2.3.1 Il magnete

Il magnete di LHCb è posizionato a circa 5 m dalla zona di interazione, ed è costituito da due bobine trapezoidali da 27 tonnellate ciascuna piegate ad un angolo di 45° nei due lati trasversali (vedi Figura 2.3). Il magnete si estende per 4.3 metri in orizzontale e 3.6 metri in verticale. Ciascuna bobina è costituita da 10 strati ricoperti da circa 3000 metri di cavi di alluminio. Il campo magnetico raggiunge un'intensità massima di circa 1 T, mentre il campo magnetico integrato $\int B dl$ lungo la sua intera lunghezza è di circa 4 T m. Lo studio delle traiettorie delle particelle deflesse dal campo magnetico permette di risalire all'impulso delle stesse, con una risoluzione $\frac{\Delta p}{p}$ dello 0.8% circa attorno a 100 GeV/c e dello 0.5% per particelle con impulso minore di 20 GeV/c. A causa della geometria non perfetta del magnete e del diverso modo di interagire col materiale del rivelatore da parte di particelle e antiparticelle, esiste una

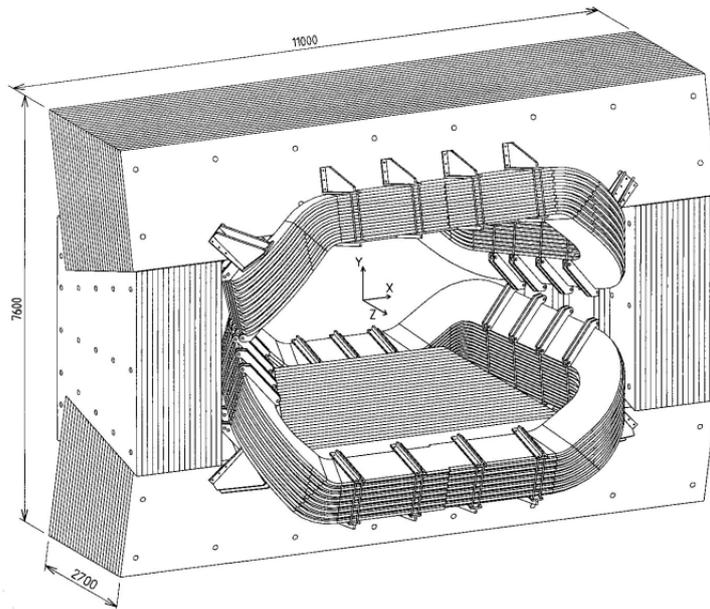


Figura 2.3: Magnete

asimmetria di rivelazione tra particelle cariche positive e particelle cariche negative. Per studiare gli effetti sistematici dovuti a questa asimmetria, durante la presa dati la polarità del magnete viene più volte invertita, passando dalla configurazione Up a Down e viceversa. Ciò è possibile grazie al fatto che il magnete non è di tipo superconduttivo, ma tradizionale.

2.3.2 VELO

Il localizzatore di vertice (VELO) circonda la zona di interazione da fasci, e misura le coordinate delle tracce lasciate dalle particelle prodotte nelle collisioni pp. Il suo scopo è quello di localizzare i vertici primari (PV), ossia i punti di interazione pp da cui hanno origine le coppie $b\bar{b}$ e $c\bar{c}$, e i vertici secondari di decadimento dei mesoni B e D. questi ultimi hanno infatti vite medie dell'ordine di 10^{-12} s e percorrono all'incirca un centimetro prima di decadere, per cui i loro punti di origine vengono ricostruiti dal VELO. Il VELO (Fig. 2.4) è costituito da 21 moduli semicircolari composti da due differenti tipi di sensori a microstrip di silicio: uno misura la coordinata radiale R, ossia la distanza dalla linea dei fasci, l'altro l'angolo azimutale ϕ , definito come l'angolo tra l'asse x e un vettore nel piano x-y nel sistema di coordinate precedentemente introdotto. I due sensori si differenziano per la segmentazione delle strip di silicio: i sensori R sono segmentati secondo cerchi concentrici, mentre i sensori ϕ sono segmentati radialmente. Entrambi i tipi di sensori hanno una maggiore densità di strip vicino ai fasci. I moduli del VELO hanno un'ampiezza di 182° , e sono disposti leggermente sfalsati in modo da sovrapporsi per evitare lacune nella accettazione.

2.3.3 Trigger Tracker

Il Trigger Tracker (TT) è posto dopo il RICH 1 e davanti all'apertura del magnete. Utilizza microstrip di silicio con un passo di $183 \mu m$, ed ha una risoluzione su una singola hit di $50 \mu m$. In Figura 2.4 è schematizzata la sua composizione. Il TT è composto da due coppie di rivelatori planari distanziate di 27 cm, di cui la prima coppia è denominata TTA e la seconda TTb. Ciascuno dei 4 strati occupa un'area di circa $2m^2$, e copre l'intera accettazione del rivelatore. Il primo e il quarto strato hanno strip

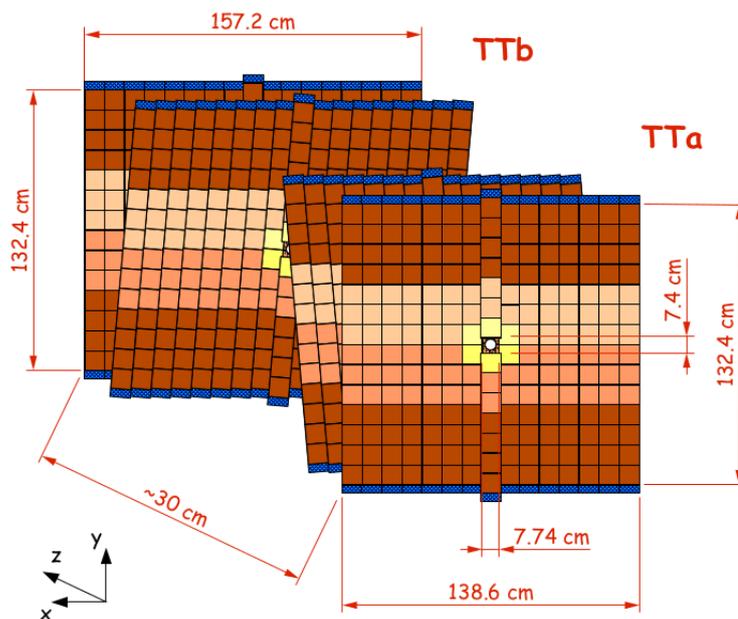


Figura 2.4: Schematizzazione del Trigger Tracker. In blu è raffigurata l'elettronica di acquisizione. Gli altri colori rappresentano i tracciatori al silicio, con le diverse tonalità a identificare i diversi passi delle strip.

disposte verticalmente, mentre il secondo ed il terzo sono inclinati di -5° e $+5^\circ$, rispettivamente. Questa configurazione permette di avere informazioni anche sulla coordinata y , oltre che sulla deflessione nel piano x - z . Il Trigger Tracker viene impiegato per misurare l'impulso trasverso delle tracce con alto parametro d'impatto e per identificare le particelle che vengono deflesse fuori dalla regione di accettazione del rivelatore e non riescono a raggiungere le stazioni T, tipicamente perché aventi impulso minore di 2 GeV/c. In maniera più generale, viene usato nella fase di trigger per verificare l'autenticità delle tracce fornite dal VELO.

2.3.4 Stazioni traccianti

Le Tracking stations rappresentano le componenti più grandi del sistema di tracciamento di LHCb e consistono di tre stazioni planari equidistanti poste dopo il magnete, denominate T1, T2 e T3. Ognuna di esse copre un'area di circa $30m^2$ ed è formata da 4 strati in configurazione analoga al TT, con il primo e quarto strato allineati verticalmente e il secondo e terzo inclinati di -5° e $+5^\circ$ rispettivamente. L'elevato flusso di particelle nella zona più vicina ai fasci rispetto alle zone più periferiche richiede che vengano utilizzate delle tecnologie più performanti nella parte centrale delle stazioni, ossia dei rivelatori a microstrip di silicio analoghi a quelli del TT, mentre la restante parte della superficie è occupata da camere a deriva, meno efficienti ma più economiche. La regione centrale del sottorivelatore prende il nome di Inner Tracker (IT), mentre quella esterna Outer Tracker (OT):

1. L'IT copre un'area di circa $125 \times 40 \text{ cm}^2$ a forma di croce attorno alla linea dei fasci, con una larghezza massima di 126 cm e una altezza massima di 41.4 cm. Come il TT è in grado di misurare la posizione di una traccia con una risoluzione di $50 \mu\text{m}$.
2. L'OT occupa la maggior parte delle stazioni traccianti. Ognuno dei dodici pannelli, di dimensioni $595 \times 480 \text{ cm}^2$, è composto da due strati di 64 straw-tubes, ossia tubi a deriva aventi diametro interno di 4.9 mm e pareti spesse $75 \mu\text{m}$. I tubi sono riempiti con una miscela gassosa di argon, anidride

carbonica (CO_2) e ossigeno (O_2) in proporzioni del 70, 28.5 e 1.5% rispettivamente. Il tempo di deriva è inferiore ai 50 ns, requisito molto importante poiché la minima separazione temporale possibile tra due pacchetti consecutivi di protoni ad LHC è di 25 ns, ed è fondamentale evitare la sovrapposizione di eventi dovuti a collisioni di protoni provenienti da 24 pacchetti differenti. Questo apparato ha un'efficienza del 99.8%, con una risoluzione sulla singola traccia di circa $210 \mu m$.

2.4 Il sistema di identificazione

I pioni costituiscono le particelle più abbondanti tra i prodotti delle collisioni pp ad LHC, mentre i decadimenti di adroni contenenti quark b e c producono numerosi kaoni, pioni, protoni e leptoni, per cui risulta necessaria la presenza di un sistema per identificare ciascuna di queste particelle per minimizzare il fondo combinatorio. Il sistema di identificazione di LHCb consiste dei due rivelatori Cherenkov, del sistema di calorimetri e delle cinque camere muoniche.

2.4.1 Rivelatori Cherenkov

I detector Ring-Imaging Cherenkov (RICH) sfruttano l'effetto Cherenkov per distinguere le tracce di kaoni, pioni e protoni. Si ha effetto Cherenkov quando una particella carica attraversa un mezzo ad una velocità β_c superiore a quella della luce in tale mezzo. In questa situazione il mezzo emette radiazione elettromagnetica, che si propaga lungo un cono di luce attorno alla direzione di propagazione con apertura angolare θ_c data da

$$\cos\theta_c = \frac{1}{n\beta} = \frac{1}{n} \frac{\sqrt{p^2c^2 + m^2c^4}}{pc}, \quad (2.1)$$

dove p è l'impulso della particella, m la sua massa a riposo ed n l'indice di rifrazione del mezzo in questione. I rivelatori RICH misurano l'angolo di Cherenkov θ_c tramite lo studio della radiazione Cherenkov; combinando il valore di θ_c con i valori di impulso ottenuti dal sistema di tracciamento si può identificare la particella in esame. In Figura 2.5 si può vedere la relazione tra l'angolo θ_c e l'impulso di varie particelle. LHCb possiede due rivelatori Cherenkov, ciascuno dei quali utilizza un materiale differente in modo da poter coprire un range di impulso da 2 a 100 GeV/c. RICH 1 è situato direttamente dopo il VELO, mentre RICH 2 è posto dopo la stazione T3. La loro composizione è visibile in Figura 2.5. Entrambi utilizzano degli specchi circolari per dirigere i fotoni sugli Hybrid Photon Detectors (HPD), ossia dei rivelatori al silicio da 1024 pixel ciascuno che rivelano il passaggio di fotoni emessi per effetto Cherenkov.

1. RICH 1 copre l'intera regione di accettazione angolare di LHCb, e riesce a discriminare le particelle con impulsi compresi tra 1 e 60 GeV/c. Per fare ciò si avvale di due mezzi differenti: aerogel con indice di rifrazione $n = 1.03$ per particelle con impulso inferiore a 9.3 GeV/c e il fluorocarburo gassoso C_4F_{10} ($n = 1.0014$) per particelle con impulso fino a 40 GeV/c.
2. RICH 2 copre una regione angolare minore di RICH 1, tra 15 e 120 mrad, poiché le particelle ad alto impulso sono prodotte prevalentemente in avanti. Il mezzo utilizzato in questo caso è tetrafluorometano gassoso, CF_4 , con indice di rifrazione $n = 1.00048$, che permette di identificare particelle con impulso compreso tra 15 e 100 GeV/c.

2.4.2 Calorimetri

Il sistema di calorimetri di LHCb consiste di quattro apparati: lo Scintillating Pad Detector (SPD), il Pre-Shower detector (PS) e i calorimetri elettromagnetico (ECAL) e adronico (HCAL). Sono tutti formati da strati di materiali assorbenti, come piombo o ferro, alternati a placche scintillatrici, e si trovano tra le camere muoniche M1 e M2. Quando una particella perde energia nel materiale assorbente si osserva la produzione di numerose particelle cariche e fotoni. Il materiale scintillatore converte tutta l'energia in fotoni, che sono poi rivelati tramite fotomoltiplicatori. Il segnale ottenuto è proporzionale all'energia iniziale della particella entrata nel calorimetro.

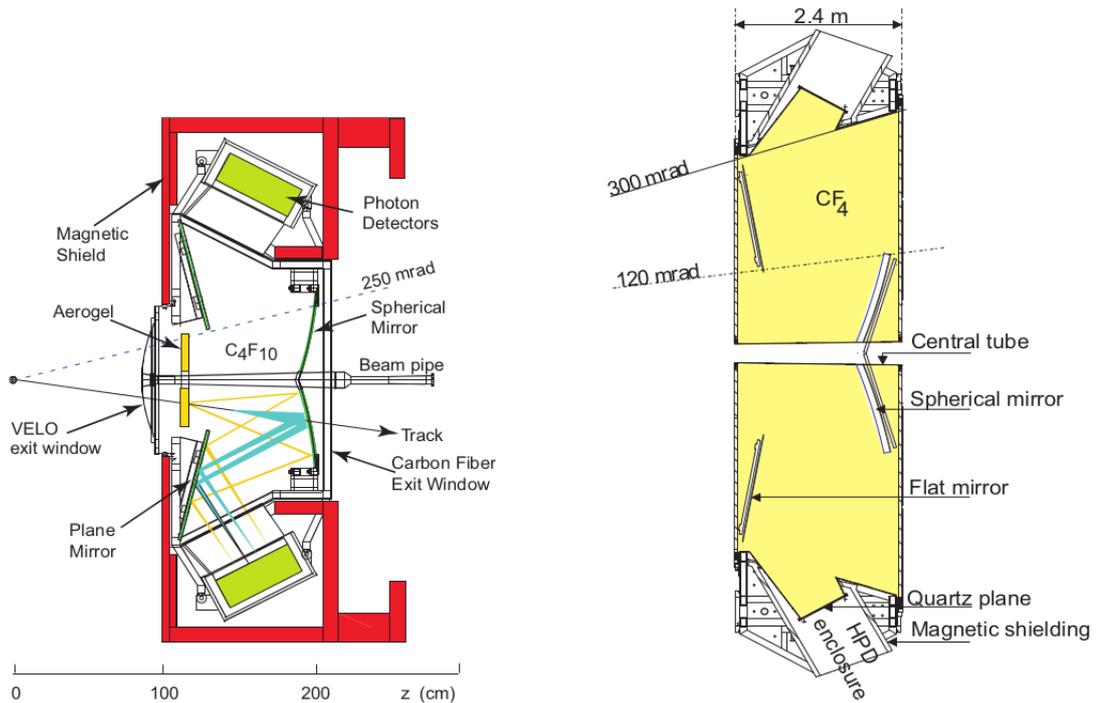


Figura 2.5: Schema dei rivelatori Cherenkov. A sinistra è disegnato il percorso dei fotoni nel rivelatore. L'asse z raffigurato corrisponde a quello del sistema di coordinate di LHCb.

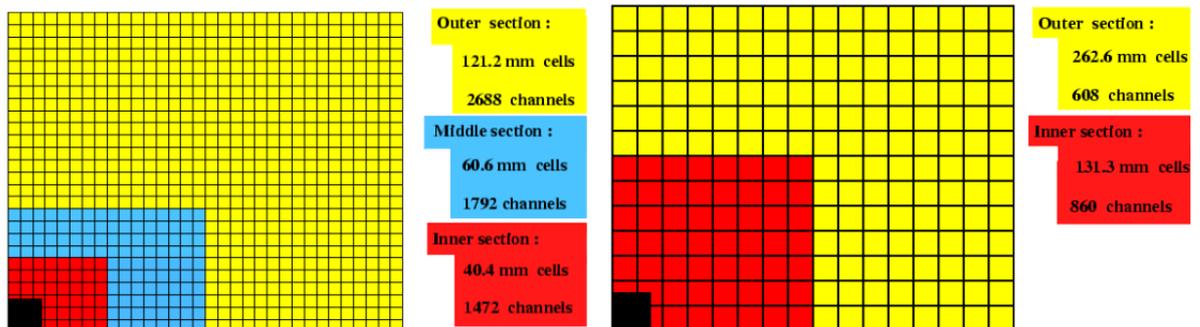


Figura 2.6: Segmentazione di un quarto della facciata anteriore di SPD, PS, ECAL (a destra) e HCAL (a sinistra).

I quattro rivelatori hanno una granulosità variabile, con celle più piccole in prossimità dei fasci e più grandi a distanze maggiori, come si può vedere in Figura 2.6. ECAL è composto da strati di piombo spessi 2 mm alternati a strati scintillatori spessi 4 mm, mentre HCAL alterna strati di ferro da 16 mm a piani di scintillatore plastico di spessore 4 mm. Le tracce rivelate da SPD, il primo della catena di calorimetri, indicano la molteplicità di particelle che raggiungono gli apparati, e dal loro numero si può decidere se accettare o rigettare gli eventi con produzione di numerose o poche particelle. ECAL e HCAL, invece, operano congiuntamente per distinguere le cascate elettromagnetiche da quelle adroniche, facilitando l'identificazione delle particelle.

2.4.3 Camere muoniche

L'identificazione dei muoni è fondamentale per ricostruire i decadimenti semileptonici dei mesoni B. Dato che i muoni hanno lunga vita media e difficilmente interagiscono con la materia, i rivelatori di muoni sono posti in fondo alla catena di detector di LHCb. Le stazioni per muoni sono cinque e consistono ciascuna di 276 camere proporzionali a multifili. Solo la prima stazione, dovendo sopportare un flusso di muoni maggiore rispetto alle altre, è equipaggiata, nella regione più vicina al fascio, con dodici rivelatori Gas Electron Multipliers (GEM), che sono in grado di sostenere un maggior flusso di particelle rispetto alle camere a fili. Le camere sono più piccole vicino al fascio e diventano sempre più grandi man mano che ci si allontana da esso. Mentre la prima stazione è posta davanti ai calorimetri, le altre quattro sono poste nella parte finale del rivelatore. Ogni camera è suddivisa in pad logici, in modo che ad ogni traccia possa essere assegnata una coordinata, permettendo così la ricostruzione della traiettoria del muone. Ogni stazione è intervallata da strati di piombo di 80 cm. Le camere sfruttano la cosiddetta geometria proiettiva, ossia la loro dimensione trasversale aumenta con la distanza dalla zona di interazione. Questo fa sì che particelle ad alto impulso, aventi una traiettoria pressoché rettilinea, colpiscano lo stesso pad in ogni stazione, facilitando la ricostruzione della loro traiettoria.

2.5 Trigger

LHCb opera ad una luminosità media di $10 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, un valore di due ordini di grandezza inferiore al limite raggiungibile da LHC. Ciò comporta due vantaggi, ossia che le componenti sono esposte ad un minore flusso di radiazioni e che ad ogni incontro dei fasci siano presenti interazioni dovute prevalentemente a collisioni singole, semplificando il compito di ricostruire le tracce delle particelle. Il sofisticato sistema di trigger di LHCb riduce l'afflusso di dati dagli iniziali 40 MHz a circa 2 kHz, frequenza alla quale i dati vengono scritti per potervi accedere nel corso di analisi successive. Questa riduzione viene ottenuta grazie a due livelli di trigger: il Livello-0 (L0), che viene eseguito a livello hardware per mezzo dell'elettronica di acquisizione, e l'High Level Trigger (HLT), che agisce a livello software da una CPU farm ed è a sua volta suddiviso in HLT1 e HLT2.

2.5.1 Livello-0

Il trigger Livello-0 riduce la frequenza di 40 MHz a cui si incrociano i fasci nel rivelatore fino a 1 MHz, frequenza alla quale si può compiere la lettura dei dati provenienti da tutte le componenti del detector. Per fare ciò compie delle stime preliminari dell'energia e dell'impulso trasversi delle particelle, e, a seconda che questi valori abbiamo superato o meno determinate soglie, decide se accettare o rigettare l'evento. Infatti i mesoni B, a causa dello loro grande massa, decadono spesso in particelle con grandi energie e impulsi trasversi. L0 si basa su tre sistemi che operano in parallelo:

1. Trigger foto/elettrico: questo trigger analizza le informazioni provenienti dai calorimetri SPD, PS e ECAL, che misurano il deposito di energia delle cascate elettromagnetiche.
2. Trigger adronico: allo stesso modo del trigger foto/elettronico, valuta l'utilità di un evento tramite i valori di energia trasversa, provenienti però dal detector HCAL, dovuti quindi a piogge adroniche.
3. Trigger muonico: utilizza le informazioni delle cinque stazioni a muoni. Le tracce vengono ricostruite definendo zone di interesse attorno alle hit delle particelle, e congiungendo quelle che giacciono nelle stesse zone. La precisione della ricostruzione dell'impulso trasverso dei muoni è dell'ordine del 25 %.

2.5.2 High Level Trigger 1

I dati vagliati da L0 vengono trasferiti in una computer farm, dove vengono elaborati dal trigger di alto livello HLT. Il primo stadio, HLT1, riduce la frequenza dei dati da 1 MHz a circa 40 kHz e compie una ricostruzione più precisa delle tracce. I criteri usati da HLT1 per decidere se un evento è interessante o meno sono derivati da alcuni vincoli particolari.

2.5.3 High Level Trigger 2

L'ultimo stadio del trigger di alto livello diminuisce la frequenza di acquisizione dei dati a 2 kHz, valore a cui è possibile scrivere i dati in maniera permanente per analisi successive. HLT2 filtra i dati analizzando le cosiddette linee topologiche del decadimento. Per prima cosa, si seleziona un insieme di tracce che soddisfano determinati requisiti di qualità della ricostruzione e di parametro d'impatto. Dopodiché si combinano due di esse a formare un oggetto a due corpi con il vincolo che la loro Distance Of Closest Approach (DOCA) sia sufficientemente piccola. Con lo stesso requisito si possono poi aggiungere fino a 2 ulteriori tracce, generando oggetti a 3 o 4 corpi. Oltre alla richiesta di prossimità delle tracce, viene aggiunto un taglio sulla massa invariante e l'impulso trasverso dell'oggetto a n corpi. Infine, viene compiuta una selezione multivariata per decidere se un candidato può essere accettato oppure no. Questo approccio permette di ricostruire con accuratezza le tracce del decadimento di un mesone B.

Capitolo 3

Metodi statistici per l'analisi dati

3.1 P.D.F. e Maximum Likelihood

Qualora si esegua un campionamento di dati di una determinata misura è possibile considerare una funzione che descriva la densità di probabilità di ogni punto nello spazio campionario; tale funzione viene detta p.d.f. (probability density function). Questa funzione dipende dall'andamento dei dati e, in particolare ha dei parametri da determinare. L'obbiettivo è quello di determinare la migliore stima dei parametri che possa rendere la p.d.f. più adatta possibile ai dati. Si consideri una variabile casuale x distribuita secondo una p.d.f. $f(x; \theta)$. Supponiamo che sia nota la forma di $f(x; \theta)$, ma che non lo sia il valore del parametro θ (o dei parametri $\theta = (\theta_1, \dots, \theta_m)$). Il metodo del Maximum Likelihood (massima verosimiglianza) è una tecnica per stimare i valori dei parametri dato un campione finito di dati. Supponiamo che una misura della variabile casuale x sia stata ripetuta n volte, ottenendo i valori x_1, \dots, x_n . Sotto l'ipotesi di $f(x; \theta)$, la probabilità che la prima misura sia nell'intervallo $[x_i, x_i + dx_i]$ è $f(x_i; \theta) dx_i$. Dal momento che le misure sono tutte ritenute indipendenti, la probabilità di avere il primo termine in $[x_1, x_1 + dx_1]$, il secondo termine in $[x_2, x_2 + dx_2]$, e così via è data da [11] :

$$P(x, \theta) = \prod_{i=1}^n f(x_i; \theta) dx_i. \quad (3.1)$$

Se la p.d.f. e i valori dei parametri sono corretti, ci si aspetta un massimo di probabilità per i dati che sono stati effettivamente misurati. Viceversa, valori dei parametri lontani dai valori reali dovrebbero produrre un valore della probabilità buono. Dal momento che il dx_i non dipende dai parametri, lo stesso ragionamento si applica anche alla seguente funzione L ,

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta), \quad (3.2)$$

chiamata funzione likelihood (o verosimiglianza). I termini dx_i sono trattati come fissati. In questo modo può determinare la miglior stima dei parametri mediante il metodo di massima verosimiglianza (ML) che consiste nel determinare i valori dei parametri che massimizzano la funzione di probabilità. Essi sono dati dalle soluzioni delle equazioni.

$$\frac{\partial L}{\partial \theta_i} = 0, i = 1, \dots, m. \quad (3.3)$$

Se esiste più di un massimo locale, verrà preso quello più alto. Indichiamo con $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)$ i valori della miglior stima dei parametri θ_i , i cui valori esatti rimangono sconosciuti.

3.2 Un esempio per ML

Supponiamo che i tempi di decadimento corretti per particelle instabili di un certo tipo siano stati misurati per n decadimenti, producendo valori t_1, \dots, t_n , e supponiamo che si scelga come p.d.f. la distribuzione di un esponenziale con media τ :

$$f(t; \tau) = \frac{1}{\tau} e^{-t/\tau}. \quad (3.4)$$

L'obbiettivo è stimare il valore del parametro τ . Al posto di utilizzare la funzione di likelihood è spesso più conveniente usare il suo logaritmo. Siccome il logaritmo è una funzione monotona crescente, il valore dei parametri che massimizza L massimizza anche $\log L$. Inoltre, il logaritmo del prodotto è la somma dei logaritmi; questa proprietà è molto utile per i calcoli, infatti [11]:

$$\log L(\tau) = \sum_{i=1}^n \log f(t_i; \tau) = \sum_{i=1}^n \left(\log \frac{1}{\tau} - \frac{t_i}{\tau} \right). \quad (3.5)$$

Massimizzando dunque la funzione di $\log L(\tau)$ si ottiene una miglior stima di τ , $\hat{\tau}$,

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n t_i \quad (3.6)$$

In questo caso $\hat{\tau}$ coincide con la media campionaria dei valori del tempo misurati. Il valore di aspettazione di $\hat{\tau}$ è

$$\begin{aligned} E[\hat{\tau}(t_1, \dots, t_n)] &= \int \dots \int \hat{\tau}(t_1, \dots, t_n) f(t_1, \dots, t_n; \tau) dt_1 \dots dt_n \\ &= \int \dots \int \left(\frac{1}{n} \sum_{i=1}^n t_i \right) \frac{1}{\tau} e^{-t_1/\tau} \dots \frac{1}{\tau} e^{-t_n/\tau} dt_1 \dots dt_n \\ &= \frac{1}{n} \sum_{i=1}^n \left(\int (t_i \frac{1}{\tau} e^{-t_i/\tau} dt_i) \prod_{j \neq i} \int \left(\frac{1}{\tau} e^{-t_j/\tau} dt_j \right) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \tau = \tau \end{aligned} \quad (3.7)$$

Concludiamo che τ è il valore di aspettazione dell'esponenziale p.d.f. e che la media campionaria è una buona stima del valore di aspettazione per qualsiasi p.d.f. .

3.3 ML con distribuzione Gaussiana: stimatori di due parametri

Supponiamo di avere misure di una variabile aleatoria x che si presume siano distribuite secondo una distribuzione gaussiana con μ e σ^2 . La funzione di log-likelihood è

$$\begin{aligned} \log L(\mu, \sigma^2) &= \sum_{i=1}^n n \log f(x_i; \mu, \sigma^2) \\ &= \sum_{i=1}^n \left(\log \frac{1}{\sqrt{2\pi}} + \frac{1}{2} \log \sigma^2 - \frac{(x_i - \mu)^2}{2\sigma^2} \right). \end{aligned} \quad (3.8)$$

Derivando $\log L$ rispetto a μ e ponendo la derivata uguale a zero e risolvendo si ottiene:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.9)$$

Calcolando il valore di aspettazione, come fatto precedentemente, si ottiene $E[\hat{\mu}] = \mu$, quindi $\hat{\mu}$ è senza bias. Ripetendo la procedura per σ^2 , si ottiene

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2. \quad (3.10)$$

calcolando il valore di aspettazione di σ^2 , si ottiene

$$E[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2 \quad (3.11)$$

Lo stimatore ML $\hat{\sigma}^2$ ha un bias diverso da zero, ma esso scompare nel limite di n grande [11].

3.4 Varianza degli stimatori ML: metodo analitico

Dato un insieme di n misure di una variabile casuale x e una ipotesi per la funzione di densità di probabilità p.d.f. $f(x; \theta)$ si è precedentemente illustrato come vengono stimati i suoi parametri. Il passo successivo è quello di dare un'incertezza statistica alle stime. Se, ad esempio, viene ripetuto l'intero esperimento un certo numero di volte (con n misurazioni ogni volta), ogni esperimento fornisce un valore di stima diverso per i parametri. Un modo per studiare la dispersione dei valori dei parametri è dato dalla varianza (o deviazione standard). Per alcuni casi si può calcolare la varianza delle miglior stime dei parametri in modo analitico. Per esempio la distribuzione esponenziale con media τ stimata da $\hat{\tau}$ ha come varianza

$$\begin{aligned} V[\hat{\tau}] &= E[\hat{\tau}^2] - (E[\hat{\tau}])^2 = \int \dots \int \left(\frac{1}{n} \sum_{i=1}^n t_i \right)^2 \frac{1}{\tau} e^{-t_1/\tau} \dots \frac{1}{\tau} e^{-t_n/\tau} dt_1 \dots dt_n \\ &- \left(\int \dots \int \left(\frac{1}{n} \sum_{i=1}^n t_i \right) \frac{1}{\tau} e^{-t_1/\tau} \dots \frac{1}{\tau} e^{-t_n/\tau} dt_1 \dots dt_n \right)^2 = \frac{\tau^2}{n}. \end{aligned} \quad (3.12)$$

Si ricordi che la varianza di $\hat{\tau}$ calcolata nell'equazione precedente è una funzione del vero (e sconosciuto) parametro di τ . Dunque cosa si prende come errore statistico dell'esperimento?

Una volta appreso il valore della varianza precedentemente calcolato, possiamo ottenere la stima ML per la varianza $\sigma_{\hat{\tau}}^2 = \tau^2/n$ semplicemente sostituendo τ con la miglior stima $\hat{\tau}$; si avrà $\hat{\sigma}_{\hat{\tau}}^2 = \hat{\tau}^2/n$, o allo stesso modo per la deviazione standard, $\hat{\sigma}_{\hat{\tau}} = \hat{\tau}/\sqrt{n}$ [11].

3.5 Metodo di Monte carlo per la varianza degli stimatori

Per casi particolarmente difficili da risolvere analiticamente, la distribuzione delle stime ML può essere studiata con il metodo di Monte Carlo. Innanzitutto vengono simulati tanti esperimenti, calcolate le stime per ognuno di essi e viene osservato come si distribuiscono i valori risultanti. Per i parametri "veri", si possono usare i valori stimati dell'esperimento reale. Dunque si può calcolare σ^2 dalle stime ML ottenute dagli esperimenti Monte Carlo e attribuire quest'ultimo come errore statistico del parametro stimato dall'esperimento reale. In molte applicazioni risulta essere troppo difficile calcolare la varianza in modo analitico e uno studio Monte Carlo di solito comporta una quantità significativa di lavoro. In questi casi si usa in genere anche la disuguaglianza Rao-Cramer-Frechet (RCF) chiamata disuguaglianza di informazioni, che dà un limite inferiore allo stimatore di varianza. Tale disequazione si applica ad ogni stimatore e non solo a quelli ottenuti dal principio di massima verosimiglianza. Per il caso del singolo parametro θ avremo:

$$V[\hat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 / E\left[-\frac{\partial^2 \log L}{\partial \theta^2}\right], \quad (3.13)$$

Dove b è il bias come definito precedentemente ed L, la funzione di likelihood. Nel caso di uguaglianza (cioè della minima varianza) lo stimatore è detto efficiente. Si può dimostrare che se esistono stimatori

efficienti per un dato problema, il metodo di massima verosimiglianza è in grado di trovarli. Inoltre può essere mostrato che gli stimatori di ML sono sempre efficienti nell'ampio limite di campionamento, tranne quando l'estensione dello spazio campionario dipende dal parametro stimato. In pratica, spesso si assume efficienza e zero bias. Per l'esempio della distribuzione esponenziale con media τ si ha:

$$\frac{\partial^2 \log L}{\partial \tau^2} = \frac{n}{\tau^2} \left(1 - \frac{2}{\tau} \frac{1}{n} \sum_{i=1}^n t_i\right) = \frac{n}{\tau} \left(1 - \frac{2\hat{\tau}}{\tau}\right) \quad (3.14)$$

e $\partial b / \partial \tau = 0$ dal momento che $b=0$. Dunque grazie al limite RCF, otteniamo che la varianza è

$$V[\hat{\tau}] \geq \frac{1}{E\left[-\frac{n}{\tau^2} \left(1 - \frac{2\hat{\tau}}{\tau}\right)\right]} = \frac{1}{-\frac{n}{\tau^2} \left(1 - \frac{2E[\hat{\tau}]}{\tau}\right)} = \frac{\tau^2}{n} \quad (3.15)$$

Dove è stata usata l'equazione per $E[\tau^2]$. Per il caso di più di un parametro, $\theta = (\theta_1, \dots, \theta_n)$ la corrispondente formula per l'inversa della matrice covariante degli stimatori $V_{i,j} = \text{cov}[\hat{\theta}_i, \hat{\theta}_j]$ è

$$(V^{-1})_{i,j} = E\left[-\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j}\right] \quad (3.16)$$

L'equazione precedente può essere riscritta come

$$(V^{-1})_{i,j} = \dots = n \cdot \int -f(x, \theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x, \theta) dx_i \quad (3.17)$$

Dove $f(x, \theta)$ è la p.d.f. per la variabile random x , per la quale si hanno n misure. L'inverso del minimo RCF per la matrice covariante è proporzionale al numero di misure nel campionamento, n . L'equazione precedente esprime il ben noto risultato che gli errori statistici (cioè la deviazione standard) decresce proporzionalmente a $1/\sqrt{n}$. In caso di un campione di dati sufficientemente ampio, si può stimare V^{-1} valutando la derivata seconda con i dati misurati e le stime ML $\hat{\theta}$:

$$(V^{-1})_{i,j} = -\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\hat{\theta}} \quad (3.18)$$

e per un singolo parametro θ :

$$\hat{\sigma}_{\hat{\theta}}^2 = \left(-1 / \frac{\partial^2 \log L}{\partial \theta^2}\right) \Big|_{\theta=\hat{\theta}} \quad (3.19)$$

Questo è il modo per stimare la matrice di covarianza quando la funzione di likelihood è massimizzata numericamente.

3.6 Metodo grafico per la varianza degli stimatori ML

Si consideri il caso di un singolo parametro θ , e espandendo la funzione "log-likelihood" in serie di Taylor con $\hat{\theta}$:

$$\log L(\hat{\theta}) = \log L(\hat{\theta}) + \left[\frac{\partial \log L}{\partial \theta}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2!} \left[\frac{\partial^2 \log L}{\partial \theta^2}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \dots \quad (3.20)$$

Per la definizione di $\hat{\theta}$, sappiamo che $\log L(\theta) = \log L_{max}$ e che il secondo termine nell'espansione è zero. Usando l'equazione 19 possiamo avere:

$$\log L(\theta) = \log L_{max} - \frac{(\theta - \hat{\theta})^2}{2\hat{\sigma}_{\hat{\theta}}^2} \quad (3.21)$$

$$\log L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) = \log L_{max} - \frac{1}{2} \quad (3.22)$$

Si può dimostrare che la funzione log-likelihood diventa una parabola nel limite di un ampio campione. Utilizzando una simulazione Monte Carlo è possibile dunque fare un grafico di $\log L(\tau)$ in funzione di τ ; la deviazione standard per $\hat{\tau}$ è stimata grazie alla formula 22, cambiando il parametro τ fino a che $\log L(\tau)$ decresce ad $1/2$. In questo modo si avrà che $\hat{\sigma}_{\hat{\tau}} \approx \Delta \hat{\tau}_- \approx \Delta \hat{\tau}_+$.

3.7 Metodo dei minimi quadrati

Il metodo dei minimi quadrati viene usato quando si vuole trovare la miglior curva che, tenendo conto degli errori, passa il più possibile vicino ai dati sperimentali determinando il best fit. Siano x_1, \dots, x_N N punti in cui si effettuano N misurazioni indipendenti y_1, \dots, y_N . I valori veri η_1, \dots, η_N delle osservabili non sono noti ma supponiamo esista un modello teorico che predica i valori η_i associati a ciascun x_i attraverso una dipendenza funzionale f , tale che:

$$\eta_i = f_i(\theta_1, \dots, \theta_L; x_i) \quad (3.23)$$

con θ_i parametri incogniti e $L \leq N$. Il principio dei minimi quadrati afferma che i valori migliori dei parametri incogniti sono quelli per cui

$$X^2 = \sum_{i=1}^N w_i (y_i - f_i)^2 \quad (3.24)$$

sia minimo, dove $w_i = \frac{1}{\sigma_{y_i}^2}$ è il peso dell' i -esima osservazione [12]. Si può dimostrare che se il modello teorico prevede una dipendenza lineare dai parametri, la minimizzazione della (24) porta ad un sistema di equazioni lineare, non omogeneo, che ammette una ed una sola soluzione esatta e non approssimata. Se le misure y_i si distribuiscono secondo una gaussiana $G(f_i, \sigma_{y_i}^2)$ e $w_i = \frac{1}{\sigma_{y_i}^2}$ allora X_{min}^2 è una variabile χ^2 con $N-L$ gradi di libertà. Il metodo dei minimi quadrati porta in questo caso alla cosiddetta minimizzazione del χ^2 e alla sua equivalenza col principio di massima verosimiglianza; infatti, $L = \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y_i - f_i}{\sigma_{y_i}})^2}$

è proporzionale a $e^{-\frac{1}{2}(\frac{y_i - f_i}{\sigma_{y_i}})^2}$ per cui L massima corrisponde al χ minimo, come si ottiene col metodo dei minimi quadrati. La funzione che associa x_i al valore calcolato teoricamente è chiaramente la p.d.f., di cui si è discusso fin'ora. La differenza principale tra il metodo dei minimi quadrati e quello della massima verosimiglianza consiste nel fatto che, col primo viene calcolato χ considerando ogni bin x_i dello spazio di campionamento ed il valore dell'osservabile corrispondente y_i viene sottratto al valore teorico della p.d.f. per quel bin. Il secondo, invece, quello di ML, non dipende dal valore dell'osservabile di ogni bin, bensì dal valore d'aspettazione. Dunque se non vi è alcun osservabile per il bin i -esimo, il metodo dei minimi quadrati fallisce, cioè il fit non converge, mentre ML funziona poichè esso non dipende dal valore osservato ma dal valore di aspettazione teorico calcolato con la p.d.f. .

3.8 Maximum likelihood con dati binnati

Consideriamo n_{tot} osservazioni di una variabile random x distribuita in accordo alla p.d.f $f(x; \theta)$ di cui vogliamo stimare il parametro sconosciuto $\theta = (\theta_1, \dots, \theta_m)$. Per un grande campione di dati diventa difficile calcolare la funzione logaritmo della Likelihood, dal momento che si dovrebbe sommare $\log(f(x_i; \theta))$ per ogni variazione del parametro θ ad ogni passo della minimizzazione. In questo caso, invece di registrare il valore di ogni misura, si può fare un istogramma, ottenendo un certo numero di entrate $n = (n_1, \dots, n_N)$ in N bins. Il valore d'aspettazione $\nu = (\nu_1, \dots, \nu_N)$ del numero di entrate sono date da [11]

$$\nu_i(\theta) = n_{tot} \int_{x_i^{min}}^{x_i^{max}} f(x; \theta) dx_i, \quad (3.25)$$

con x_i^{min} e x_i^{max} limiti del bin. Si può considerare l'istogramma come una singola misura di un N -dimensionale vettore random per cui la p.d.f. totale è data da una distribuzione del tipo

$$f(n; \nu) = \frac{n_{tot}!}{n_1! \dots n_N!} \left(\frac{\nu_1}{n_{tot}}\right)^{n_1} \dots \left(\frac{\nu_N}{n_{tot}}\right)^{n_N}. \quad (3.26)$$

La probabilità di essere nel bin i è espressa come valore di aspettazione diviso per il numero totale di osservazioni n_{tot} . Prendendo il logaritmo della funzione

$$\log(L(\theta)) = \sum_{i=1}^N n_i \log(\nu_i(\theta)), \quad (3.27)$$

Le miglior stime $\hat{\theta}$ sono trovate massimizzando $\log(L)$ da qualunque media disponibile. Nel limite che la grandezza del bin è molto piccola (cioè N molto grande) la funzione di likelihood diventa la stessa di quella del metodo di ML senza binnaggio. Perciò non si hanno problemi se alcuni bins hanno poche entrate oppure nessuna, diversamente dal metodo dei minimi quadrati.

3.9 La funzione di Maximum Likelihood Estesa

Consideriamo una variabile random x distribuita secondo una p.d.f. $f(x;\theta)$ con parametri $\theta=(\theta_1, \dots, \theta_m)$, e supponiamo di avere un campione di dati x_1, \dots, x_n . C'è spesso il caso in cui il numero di osservazioni n nel campionamento è esso stesso una variabile random di Poisson con valore medio ν . La funzione di Likelihood è il prodotto della probabilità di Poisson di trovare n , e la funzione di likelihood per gli n valori di x [11],

$$L(\nu, \theta) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^n f(x_i; \theta) = \frac{e^{-\nu}}{n!} \prod_{i=1}^n \nu f(x_i; \theta). \quad (3.28)$$

Questa è chiamata funzione di Likelihood estesa. Si possono distinguere due situazioni: una in cui ν è trattata come una funzione di θ , l'altra in cui sia un parametro indipendente.

1. Assumiamo che ν sia una funzione di θ . La funzione logaritmo di Likelihood estesa è

$$\log L(\theta) = n \log(\nu(\theta)) - \nu(\theta) + \sum_{i=1}^n \log(f(x_i, \theta)) = -\nu(\theta) + \sum_{i=1}^n \log(\nu(\theta) f(x_i; \theta)), \quad (3.29)$$

Includendo il termine di Poisson avremo una varianza minore per la miglior stima $\hat{\theta}$ di quella che avremmo se usassimo solo x come variabile. La deviazione standard della miglior stima corrisponde all'ammontare delle stime che fluttuano se uno ripete l'esperimento tante volte ed ogni volta con diverso numero di eventi.

2. La seconda situazione si ha quando non c'è una relazione funzionale tra ν e θ . Prendendo il logaritmo di 3.28 e ponendo la sua derivata rispetto a ν uguale a zero otteniamo lo stimatore

$$\hat{\nu} = n \quad (3.30)$$

Derivando la funzione $\log L(\nu, \theta)$ rispetto a θ_i e ponendola uguale a zero, si ottengono le stesse stime $\hat{\theta}$ del caso di ML. Dunque la situazione è identica a quella di prima, solo che ora una quantità che dipende sia da n che da $\hat{\theta}$ contiene una fonte di fluttuazione statistica, dal momento che n è una variabile random.

Consideriamo, ora, che la p.d.f. di una variabile x sia la sovrapposizione di tante componenti, cioè la somma

$$f(x, \beta_i) = \sum_{i=1}^m \beta_i f_i(x), \quad (3.31)$$

e vogliamo stimare i β_i che rappresenta la contribuzione di ogni componente. Se le funzioni di probabilità sono note, si possono considerare i parametri β_i non tutti indipendenti. Nel caso solito senza la funzione di Likelihood estesa si possono realizzare i parametri, sostituendo uno dei coefficienti delle funzioni, per

esempio β_m , con $1 - \sum_{i=1}^{m-1} \beta_i$, così da avere solo $m-1$ parametri. Si può trattare lo stesso problema con la ML estesa. Prendendo il logaritmo della 3.28 (?) e trascurando i termini non dipendenti dai parametri si ottiene

$$\log(L(\nu, \beta)) = -\nu + \sum_{i=1}^n \log\left(\sum_{j=1}^m \nu \beta_j f_j(x_i)\right). \quad (3.32)$$

Definendo $\mu = \beta_i \nu$ come valore d'aspettazione del numero di eventi di tipo i , la funzione logaritmo di Likelihood può essere scritta come una funzione degli m parametri $\mu = (\mu_1, \dots, \mu_m)$,

$$\log(L(\nu, \beta)) = -\sum_{j=1}^m \mu_j + \sum_{i=1}^n \log\left(\sum_{j=1}^m \mu_j f_j(x_i)\right). \quad (3.33)$$

I parametri μ non sono soggetti a nessun vincolo, come invece lo sono le componenti di β . Il numero totale di eventi n è visto ora come la somma di variabili indipendenti di Poisson con media μ_i . In questo modo tutti i parametri μ_i vengono trattati come variabili poissoniane e quindi simmetricamente.

Capitolo 4

Risultati per le simulazioni di Monte Carlo

4.1 Modello teorico e parametri della P.D.F. nella simulazione

Come esempio di quanto detto nel paragrafo 3.1.9. si consideri un campionamento di dati di due tipi di eventi, segnale e background, dove ogni evento è caratterizzato da una variabile continua x . Supponiamo che gli eventi di segnale siano descritti da $f_s(x)$, una distribuzione gaussiana, e il background da $f_b(x)$, una distribuzione esponenziale. Il numero di eventi di segnale n_s è distribuito secondo una distribuzione poissoniana con media μ_s , e il numero di eventi del background n_b è distribuito secondo una distribuzione poissoniana μ_b . Dunque la p.d.f. per x è

$$f(x) = \frac{\mu_s}{\mu_s + \mu_b} f_s(x) + \frac{\mu_b}{\mu_s + \mu_b} f_b(x), \quad (4.1)$$

dove le funzioni $f_s(x)$ e $f_b(x)$ sono note, e sono normalizzate ad uno. f_s dipende dai parametri μ e σ e dalla variabile x , $f_b(x)$ invece dal parametro λ e dalla variabile x . ognuna delle due funzioni è moltiplicata per il relativo coefficiente s o b , che sono anch'essi dei parametri da stimare. Tali parametri rappresenteranno i coefficienti della 3.34 nella sezione precedente e in particolare avremo che:

$$s = \frac{\mu_s}{\mu_s + \mu_b}, b = \frac{\mu_b}{\mu_s + \mu_b} \quad (4.2)$$

dove μ_s e μ_b sono la media della distribuzione poissoniana del segnale (n_s) e la media del background (n_b), rispettivamente.

4.2 Simulazione dati

Fino ad ora abbiamo parlato di distribuzione di probabilità di una generica variabile x . Il nostro scopo però è quello di valutare quale tra i metodi possibili è il migliore per fare il fit della massa invariante nel decadimento del D^0 . A questo proposito consideriamo come variabile della nostra distribuzione di probabilità la massa invariante del D^0 . In una collisione protone-protone genera una quantità di mesoni B e D molto grande in una collisione; a loro volta tali particelle decadono in altri mesoni. In particolare gli eventi ricostruiti dal rivelatore LHCb del decadimento della particella del D^0 possono essere decine di milioni. Per questo nella simulazione fatta si ritiene più appropriato l'utilizzo di una statistica alta. Si sono infatti generati 50 milioni di eventi come numero totale di generazioni. Tali eventi saranno dunque suddivisi in eventi di segnale n_s ed eventi di fondo n_b . Si stima che i due tipi di eventi durante la presa dati possano essere in egual numero: 50 % n_s e 50 % n_b . Ovviamente ci sono tanti fattori che influenzano le frazioni di segnale e fondo ma la situazione descritta è verosimile. Inoltre al fine della simulazione

è irrilevante considerare diverse frazioni per cui si è scelto un fondo grande quanto il segnale. Si può dunque considerare come p.d.f. per la massa invariante la sovrapposizione descritta nella sezione 4.1 con segnale gaussiano e rumore di fondo descritto da un esponenziale. E' possibile a questo punto simulare una serie di dati e fare il fit di essi utilizzando i due metodi descritti precedentemente. In realtà, come abbiamo visto, il metodo del Maximum Likelihood può essere utilizzato con o senza binnaggio, mentre per il motivo espresso precedentemente il metodo dei minimi quadrati richiede necessariamente l'uso del binnaggio. Dunque per un fatto di distinzione usiamo L per indicare il maximum likelihood mentre \bar{L} per indicare il Maximum Likelihood con l'uso del binnaggio. Siccome analizziamo la massa invariante del D^0 , vogliamo determinare μ , σ ed n_s con gli stimatori $\hat{\mu}$, $\hat{\sigma}$ e \hat{n}_s coi diversi metodi, sapendo che per una particella D^0 abbiamo una massa $M_{D^0} = (1864 \pm 10) MeV$. Scegliamo dunque una statistica di 50 milioni di eventi ed un range di massa di 120 MeV, simmetrico al valore medio. Per la simulazione si è scelto di utilizzare il programma Root e in particolare si è usata la libreria RooFit [?]. La scelta di questo software ha ottimizzato e accelerato il calcolo e lo studio dei parametri dei fit per i metodi statistici usati. Root nasce proprio come programma finalizzato all'analisi e alla rappresentazione di dati o simulazione dati. La simulazione è stata fatta variando il numero totale di bin; si sono infatti utilizzati 5, 10, 50, 100, 200, 300, 500 e 1000 bin. In la Tabella 2. sono riportati i risultati ottenuti per la simulazione, dove in particolare per ogni configurazione sono riportati i valori dei parametri ottenuti dall'algoritmo di minimizzazione.

Bins		$\hat{\mu}(MeV)$	$\hat{\sigma}(MeV)$	$s(x 10^6)$	$bs(x 10^6)$
5	χ^2	1864.010 ± 0.003	12.314 ± 0.003	24.7599 ± 0.0083	25.0259 ± 0.0069
	\bar{L}	1864.010 ± 0.003	10.603 ± 0.003	25.0196 ± 0.0070	24.9954 ± 0.0070
10	χ^2	1864.010 ± 0.003	10.001 ± 0.002	25.0019 ± 0.0068	24.9790 ± 0.0069
	\bar{L}	1864.000 ± 0.003	9.9951 ± 0.003	24.9856 ± 0.0069	24.9889 ± 0.0068
50	χ^2	1864.010 ± 0.003	10.026 ± 0.003	25.0046 ± 0.0068	24.9960 ± 0.0069
	\bar{L}	1864.000 ± 0.003	10.000 ± 0.003	25.0104 ± 0.0068	24.9897 ± 0.0069
100	χ^2	$1864.010 \pm 0,003$	$10.009 \pm 0,003$	25.0044 ± 0.0069	24.9957 ± 0.0069
	\bar{L}	1864.000 ± 0.003	10.001 ± 0.003	25.0080 ± 0.0068	25.0040 ± 0.0069
200	χ^2	1864.010 ± 0.003	10.004 ± 0.003	25.0043 ± 0.0069	24.9956 ± 0.0069
	\bar{L}	1864.000 ± 0.003	10.001 ± 0.003	24.9891 ± 0.0069	25.0077 ± 0.0068
300	χ^2	1864.010 ± 0.003	10.003 ± 0.003	25.0044 ± 0.0068	24.9953 ± 0.0069
	\bar{L}	1864.000 ± 0.003	$9.997 \pm 0,003$	24.9923 ± 0.0068	25.0077 ± 0.0068
500	χ^2	1864.010 ± 0.003	10.003 ± 0.003	25.0044 ± 0.0068	24.9952 ± 0.0068
	\bar{L}	1864.000 ± 0.003	$10,000 \pm 0.003$	24.9935 ± 0.0068	25.0065 ± 0.006858
1000	χ^2	1864.01 ± 0.00275615	10.0025 ± 0.00262498	25.0044 ± 0.0068	24.9947 ± 0.0068
	\bar{L}	$1864.000 \pm 0,003$	9.999 ± 0.003	24.9925 ± 0.0068	25.0001 ± 0.0069
	L	1864.010 ± 0.003	10.000 ± 0.004	25.0004 ± 0.0069	25.0001 ± 0.0068

Tabella 2. Risultati dei parametri ottenuti dalla simulazione per i metodi di ML binnato e del χ^2 a confronto e il risultato della simulazione per il metodo ML

Di seguito riportiamo i grafici delle simulazioni di χ^2 ed ML binnato e mettiamo a confronto i grafici per lo stesso numero di bins. In Figura 4.1 si mettono a confronto i fit di ML binnato e quello di χ^2

usando 5 bins. In Figura 4.2 si mettono a confronto i fit di ML binnato e quello di χ^2 usando 10 bins. In Figura 4.3 si mettono a confronto i fit di ML binnato e quello di χ^2 usando 50 bins. In Figura 4.4 si mettono a confronto i fit di ML binnato e quello di χ^2 usando 100 bins. In Figura 4.5 si mettono a confronto i fit di ML binnato e quello di χ^2 usando 200 bins. In Figura 4.6 si mettono a confronto i fit di ML binnato e quello di χ^2 usando 300 bins. In Figura 4.7 si mettono a confronto i fit di ML binnato e quello di χ^2 usando 500 bins. In Figura 4.8 si mettono a confronto i fit di ML binnato e quello di χ^2 usando 1000 bins. In Figura 4.9 si osserva il fit con il metodo del ML non binnato.

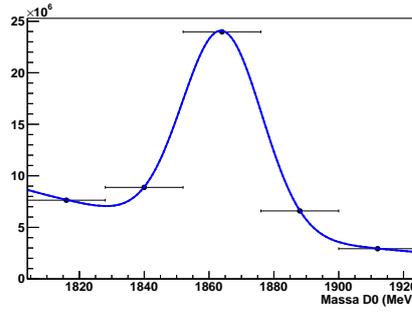


Figura 4.1: Fit con 5 bins con χ^2 ; col metodo del ML non si ha un fit convergente

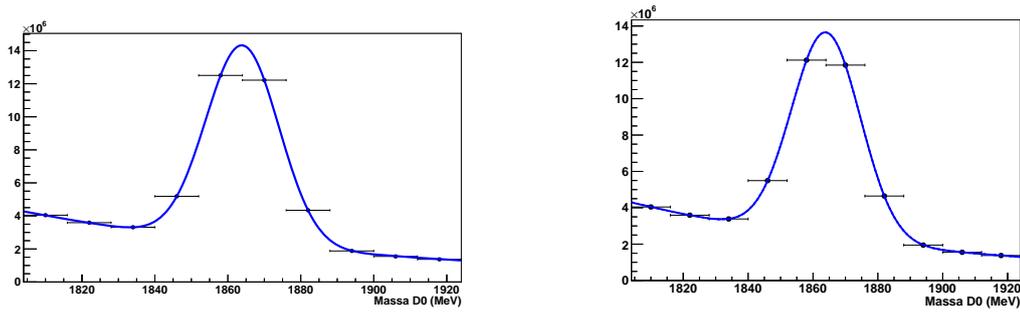


Figura 4.2: Fit con 50 bins: a sinistra con ML e a destra con χ^2

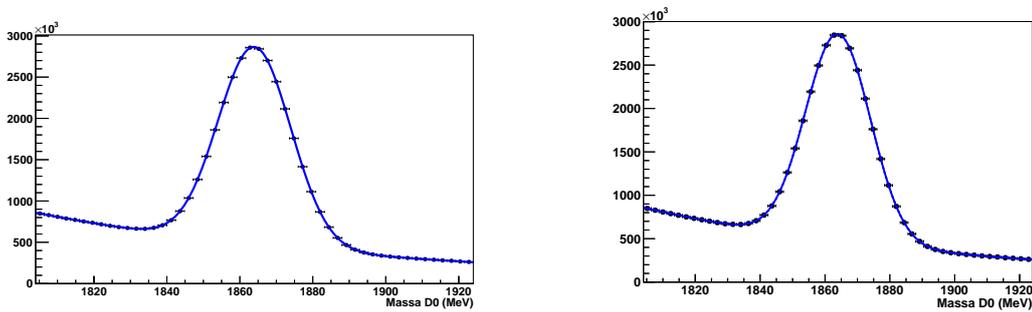


Figura 4.3: Fit con 50 bins: a sinistra con ML e a destra con χ^2

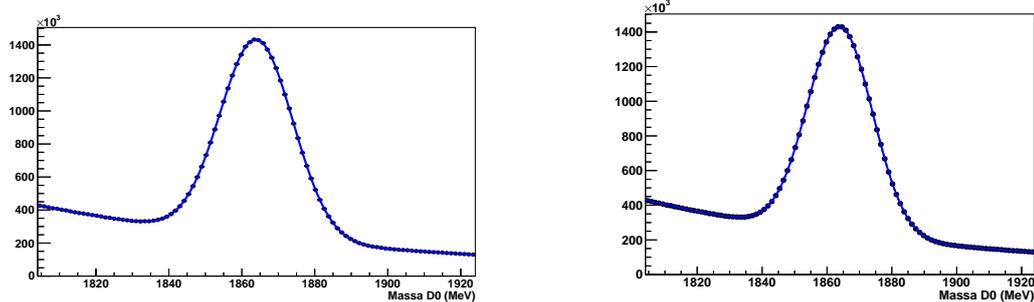


Figura 4.4: Fit con 100 bins: a sinistra con ML e a destra con χ^2

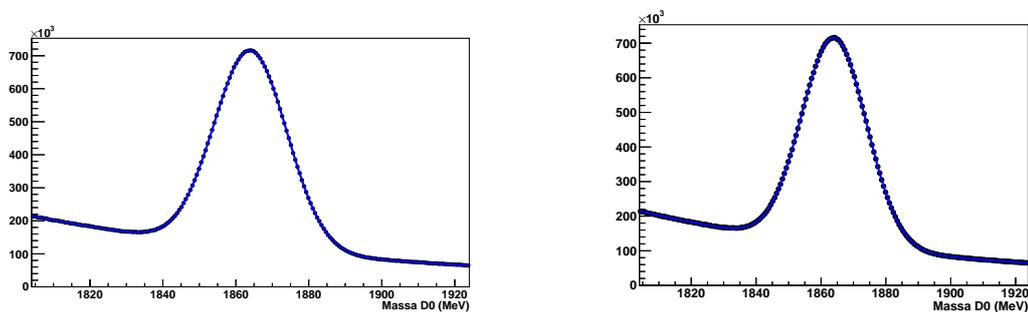


Figura 4.5: Fit con 200 bins: a sinistra con ML e a destra con χ^2

Confrontando i risultati della Tabella 2, si osserva come le migliori stime calcolate per la maggior parte dei casi siano compatibili con i valori veri di μ e σ forniti per la simulazione; non vi è alcuna predilezione di un metodo piuttosto che l'altro, perché con la p.d.f. utilizzata la media e la deviazione standard tendono con la minimizzazione (dei quadrati) o massimizzazione (della funzione di verosimiglianza) ad individuare lo stesso valore di massa. In realtà non sempre si possono usare tali metodi; se si prende, infatti, $n_{bin} = 5$, si nota dai dati della simulazione (Tabella 2) che il valore della deviazione standard del parametro ottenuto con il metodo del χ^2 ha una distanza di più di 2 MeV dal valore vero fornito con la simulazione. Dunque per la simulazione fatta è più opportuno utilizzare un numero di bins maggiore di 10; da questo valore, infatti, sempre riferendosi alla Tabella 1 si ha un accordo sempre più buono tra i parametri e i valori veri.

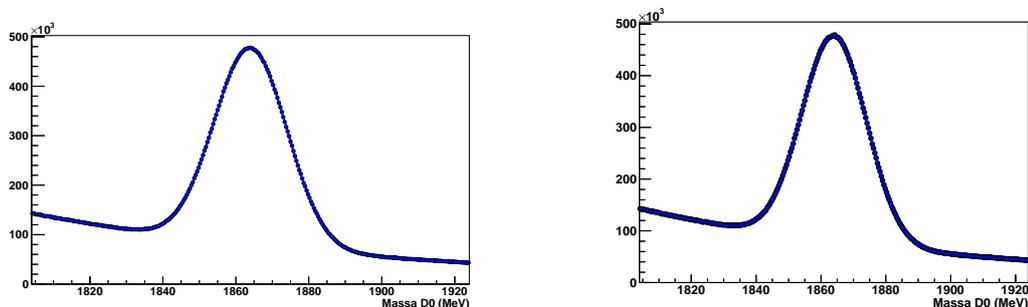


Figura 4.6: Fit con 300 bins: a sinistra con ML e a destra con χ^2

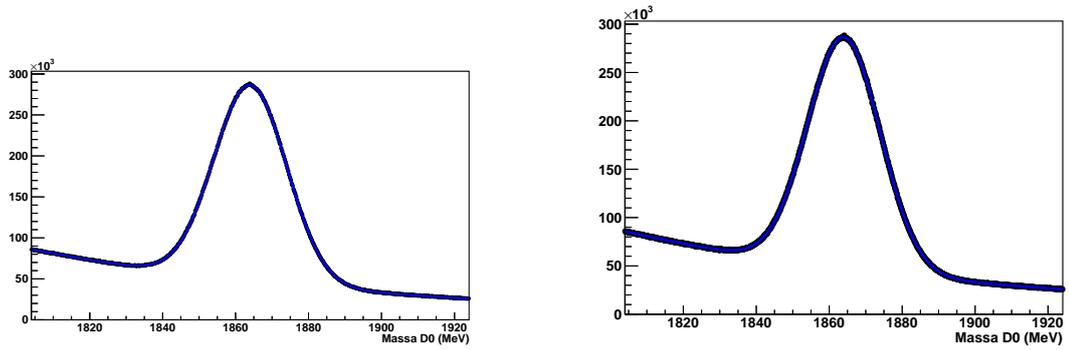


Figura 4.7: Fit con 500 bins: a sinistra con ML e a destra con χ^2

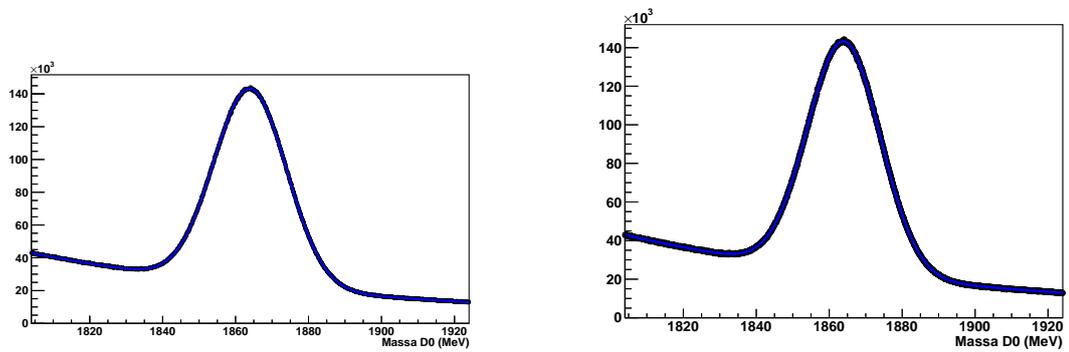


Figura 4.8: Fit con 1000 bins: a sinistra con ML e a destra con χ^2

Un'altro problema consiste quando si ha una statistica bassa, ma un numero di bin alto; si trova, infatti, secondo queste condizioni, che per l'*i*-esimo bin può non presentarsi alcun evento; non è possibile dunque utilizzare il metodo del χ^2 o dei minimi quadrati in questo caso perchè per usare quest'ultimo è necessario avere un numero abbastanza alto di entrate per bin. Siccome per il decadimento del D^0 si hanno milioni di eventi non sarà mai necessario usare una bassa statistica per cui è sempre applicabile il metodo del

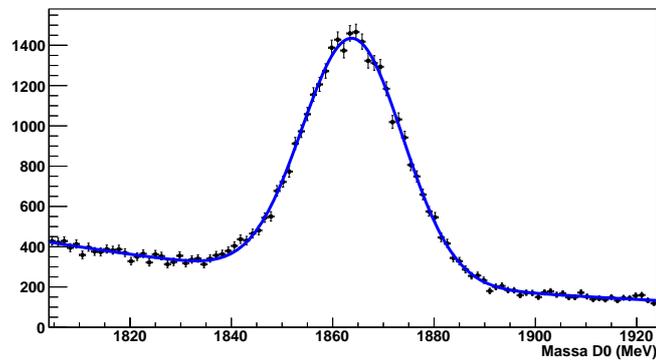


Figura 4.9: Metodo ML non binnato con 50 milioni di eventi

χ^2 nelle nostre condizioni. Come detto precedentemente a livello di risultati non vi è una discrepanza notevole tale per cui si possa scegliere un metodo piuttosto che un altro, ma per statistiche elevate può sorgere un problema: calcolando, infatti, il tempo di elaborazione del fit nei tre casi si ha una notevole differenza. Quando si usa il metodo del χ^2 o quello di ML binnato si ha un “tempo di fit” relativamente basso, mentre questo aumenta considerevolmente utilizzando il metodo ML, sempre con statistiche alte. Di seguito è mostrata la Tabella 3 che mostra il ”tempo di fit“ per 50 milioni di eventi nei tre casi.

	Tempo di fit
χ^2	9.5 s
ML binnato	9.1 s
ML	8.5×10^3 s

Tabella 3. Valori del tempo di esecuzione del fit da parte del Programma Roofit con i diversi metodi.

Notiamo che seppure ci sia una differenza di circa 0.4 secondi tra il metodo del χ^2 e quello del ML con distribuzione binnata, comunque essa non risulta rilevante (il tempo riportato in tabella risulta essere il tempo medio delle 8 simulazioni). E' invece piuttosto rilevante la differenza tra i due metodi sopracitati e quello del ML con distribuzione non binnata; quest'ultimo ha un ”tempo di fit” molto elevato di circa 2 ore 23 minuti e 4,63 secondi. Dunque per statistiche alte è più conveniente utilizzare i metodi che ottimizzano il tempo: quello del χ^2 e quello di ML binnato.

Conclusioni

Mediante simulazioni Monte Carlo di distribuzioni di massa invariante del mesone D^0 si è arrivati alla considerazione che per il calcolo dei parametri del fit sui dati simulati è indifferente, considerando un'alta statistica la scelta di usare il metodo del maximum likelihood piuttosto che quello dei minimi quadrati, purchè si adottino una distribuzione di massa binnata con un numero di bin sufficientemente grandi. Si è visto infatti che con l'uso del metodo di ML con distribuzione non binnata si ha un tempo di esecuzione del fit, da parte del programma di Root, di circa 8.5×10^3 secondi, mentre per quanto riguarda i metodi con la distribuzione binnata del χ^2 e del ML si ha un tempo di fit di pochi secondi, con una differenza irrilevante. Si è notato, inoltre, che ci sono dei limiti nell'utilizzo dei due metodi con distribuzione binnata. Se si prende, infatti, un numero di eventi abbastanza basso, il metodo del χ^2 è inutilizzabile, in quanto si hanno pochi eventi per bin o addirittura nessun evento per determinati bin. Questo problema, in realtà non sussiste, in quanto per l'analisi della massa invariante del D^0 ad LHCb si hanno decine di milioni di eventi; per questo motivo si è deciso di utilizzare nella simulazione una statistica di 50×10^6 eventi. Un'altro limite di utilizzo si ha se viene considerato, viceversa, un numero di bin basso; con l'uso di 5 o 10 bin infatti si ottengono dei parametri del fit abbastanza distanti da quelli nominali utilizzati da una fase di simulazione. Se, invece, si utilizzano 50, 100, 200, 300, 500 o 1000 bin si ha che la miglior stima dei parametri $\hat{\mu}$, $\hat{\sigma}$ e \hat{n}_s è compatibile con i valori nominali dei parametri generati. Dunque i metodi con distribuzione binnata sono da rigettare nel caso si usi un numero di bin basso, mentre risultano adeguati quando viene usato un numero di bin abbastanza elevato, superiore a 50.

Bibliografia

- [1] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, *Evidence for the 2π Decay of the K_S^0 Meson*, Phys. Rev. Lett. **13** (1964) 138.
- [2] N. S. cesari (dispense di lezione del corso di introduzione alla fisica nucleare e subnucleare), *Fisica nucleare e subnucleare*, .
- [3] B. Martin, *Nuclear and particle physics, an introduction*, 2006 John Wiley Sons, Ltd. ISBN 0-470-01999-9.
- [4] M. Gersabeck, *Introduction to Charm Physics*, PoS **FWNP** (2015) 001, arXiv:1503.0003.
- [5] CDF collaboration, T. Aaltonen *et al.*, *Measurement of the difference of CP-violating asymmetries in $D^0 \rightarrow K^+K^-$ and $D^0 \rightarrow \pi^+\pi^-$ decays at CDF*, Phys. Rev. Lett. **109** (2012) 111801, arXiv:1207.2158.
- [6] BaBar collaboration, B. Aubert *et al.*, *Search for CP violation in the decays $D^0 \rightarrow K^-K^+$ and $D^0 \rightarrow \pi^-\pi^+$* , Phys. Rev. Lett. **100** (2008) 061803, arXiv:0709.2715.
- [7] Belle collaboration, M. Staric *et al.*, *Measurement of CP asymmetry in Cabibbo suppressed D^0 decays*, Phys. Lett. **B670** (2008) 190, arXiv:0807.0148.
- [8] B. R. Ko, *CP violation and mixing in the charm sector at Belle, and current HFAG averages*, arXiv:1212.5320.
- [9] LHCb collaboration, R. Aaij *et al.*, *Measurement of CP asymmetry in $D^0 \rightarrow K^-K^+$ and $D^0 \rightarrow \pi^-\pi^+$ decays*, JHEP **07** (2014) 041 CERN-PH-EP-2014-082, LHCb-PAPER-2014-013, arXiv:1405.2797.
- [10] LHCb collaboration, R. Aaij *et al.*, *Measurement of the difference of time-integrated CP asymmetries in $D^0 \rightarrow K^-K^+$ and $D^0 \rightarrow \pi^-\pi^+$ decays*, Phys. Rev. Lett. **116** (2016) 191601 LHCb-PAPER-2015-055, CERN-EP-2016-022, arXiv:1602.0316.
- [11] G. Cowan, *Statistical data analysis*, Oxford University Press Inc. , New York.
- [12] J. Taylor-Zanichelli, *Introduzione all'analisi degli errori, lo studio delle incertezze nelle misure fisiche*, .