

Metodi statistici per la fisica delle particelle elementari:

Appunti dalle lezioni del Prof. Riccardo Faccini ¹

Dott. R. Faccini, A. D'Orazio, S. Fiore, Università degli studi La Sapienza di Roma

28 luglio 2015

¹nell'ambito del corso di Laboratorio di Fisica Nucleare e Subnucleare del prof. Fernando Ferroni nel corso di laurea in fisica.

Indice

1	Introduzione	5
1.1	Variabili casuali e funzioni di distribuzione	5
1.1.1	<i>pdf</i> a più variabili	6
1.1.2	Funzione di variabili casuali	6
1.2	Proprietà della <i>pdf</i>	7
1.2.1	<i>moda</i>	7
1.2.2	<i>mediana</i>	7
1.2.3	Valore di aspettazione	7
1.2.4	Varianza	8
1.3	Funzione di distribuzione cumulativa	10
1.4	Estimatori	10
1.5	Campioni di dati con stessa <i>pdf</i>	11
1.6	Esercizi	12
2	Funzioni di distribuzione	13
2.1	Distribuzione binomiale	13
2.1.1	Applicazione: l'efficienza di selezione	15
2.2	Distribuzione di Poisson	15
2.2.1	Somma di variabili poissoniane	17
2.2.2	Applicazione: l'istogramma	17
2.3	Distribuzione uniforme	18
2.3.1	Applicazione: packing dei dati.	19
2.4	Distribuzione normale o gaussiana	19
2.4.1	Proprietà distribuzione Gaussiana	20
2.4.2	<i>pdf</i> Gaussiana a più variabili	20
2.5	Teorema del limite centrale	22
2.6	Esercizi	22
3	Statistica di Bayes	25
3.1	Teorema di Bayes	25
3.1.1	Applicazione alle misure	27
3.1.2	Esempio: misura di vita media	27
3.2	Esercizi	29
4	Tecniche di fit	31
4.1	Verosimiglianza (<i>likelihood</i>)	31
4.1.1	Caratteristiche necessarie della <i>likelihood</i>	31
4.1.2	Estimatori ottenuti con la <i>likelihood</i>	32

4.1.3	Likelihood estesa	36
4.1.4	Esempi	36
4.2	χ^2 e minimi quadrati	39
4.2.1	Media pesata	39
4.2.2	Esempio: Fit dei dati ai parametri del Modello Standard	41
4.3	Fit ad istogrammi	41
4.4	Esercizi	41
5	Test di ipotesi	43
5.1	Distribuzione del chiquadro: χ^2	44
5.1.1	Esempi	45
5.2	Test di student	46
5.3	Test di Kolmogorof	46
5.4	Esercizi	46
6	Metodi MonteCarlo	47
6.1	Metodi Montecarlo	47
6.1.1	Generazione di numeri casuali uniformi	47
6.1.2	Metodo di Wilson	47
6.1.3	Metodo diretto	47
6.1.4	Hit & Miss	47
6.1.5	Campionamento di importanza	47
6.1.6	Metodo dei pesi	47
6.2	Esempio in fisica delle alte energie	47
7	Estrazione di limiti	49
7.1	Conteggio di eventi	49
7.1.1	limiti in assenza di fondo	49
7.1.2	limiti in presenza di fondo	49
7.2	limiti con ikelihood	49
7.3	Feldman Cousin	49
7.4	Esercizi	49
8	Soluzione degli esercizi	51

Capitolo 1

Introduzione

Scopo della sperimentazione in fisica è trovare modelli interpretativi dei fenomeni fisici osservati. Vengono perciò sviluppati modelli teorici, che generalmente dipendono da parametri incogniti, e vengono poi realizzati esperimenti per verificare la correttezza dei modelli e misurarne i parametri.

In questo procedimento sono necessarie tecniche di analisi statistica :

- per verificare se i dati sono in accordo con una determinata ipotesi (*Test di Ipotesi*). Varie sono le possibili applicazioni, dalla verifica che un modello riproduca correttamente i dati sperimentali, alla verifica che più esperimenti siano compatibili tra loro.
- per estrarre dai dati *estimatori* dei parametri dei modelli.

Questo corso si ripropone di insegnare le tecniche più comunemente usate tramite esempi specifici per la fisica delle particelle elementari ed esercizi. Una trattazione più rigorosa di questi argomenti si può trovare, per esempio, in [1, ?, ?].

1.1 Variabili casuali e funzioni di distribuzione

Il risultato di un esperimento è una variabile casuale (VC), cioè un numero reale che fa parte di un insieme di numeri la scelta tra i quali non è sotto il controllo dello sperimentatore. Ogni qual volta l'esperimento viene ripetuto la variabile casuale assume un valore diverso all'interno dell'insieme permesso. La probabilità di ottenere un dato valore, cioè la legge che regola la scelta della variabile all'interno dell'insieme di cui fa parte, è detta *distribuzione di probabilità* (probability density function - *pdf*). Il legame tra esperimento e modello avviene dunque tramite la *pdf*, predetta dal modello per descrivere l'esperimento.

Se la VC può prendere solo un numero discreto di valori, questa si definisce *discreta*, altrimenti *continua*.

Nel caso di variabili discrete, la probabilità che una variabile casuale X assuma il valore x_i è

$$P(X = x_i) = P_i \tag{1.1}$$

La condizione di normalizzazione, dati tutti i possibili valori i che una VC può assumere, è

$$\sum_i P_i = 1 \quad (1.2)$$

Nel caso di variabili che assumono un continuo di valori si introduce la probabilità che la VC assuma un valore nell'intervallo infinitesimo

$$P(x \leq X \leq x + dx) = f(x)dx \quad (1.3)$$

con

$$\int_{\Omega} f(x)dx = 1 \quad (1.4)$$

dove Ω è l'intervallo di valori assumibili (*spazio di campionamento*).

1.1.1 pdf a più variabili

Quando si misurano più variabili casuali allo stesso tempo, la forma più generale della pdf è una generica funzione di tutte le variabili $f(\vec{x})$ il cui significato è che in un cubo infinitesimo centrato in \vec{x} , $I_{\vec{x}, d\vec{x}}$ di lati dx_1, \dots, dx_N

$$P(X \in I_{\vec{x}, d\vec{x}}) = f(\vec{x})d\vec{x} \quad (1.5)$$

Tutte le espressioni possono essere tradotte da una a molte variabili semplicemente cambiando lo spazio di integrazione.

Caso particolare è quello in cui le variabili sono tra loro scorrelate, cioè il comportamento di ognuna di esse non dipende dal valore delle altre. In questo caso è possibile definire la pdf (f_i) per ognuna variabile separatamente:

$$f(\vec{x}) = \prod_i f_i(x_i) \quad (1.6)$$

1.1.2 Funzione di variabili casuali

La variabile misurabile potrebbe essere una funzione delle variabili casuali di cui si conosce la pdf: $g = g(\vec{x})$ (oppure $\nu = \nu(\vec{N})$ nel caso di variabili discrete). In questi casi è spesso necessario calcolare la pdf della variabile $g(\nu)$, sommando su tutti i valori delle variabili casuali che forniscono un dato valore della variabile di interesse:

- in caso di variabili discrete

$$f'(\nu) = \sum_{\vec{N}} \delta_{\nu, \nu(\vec{N})} f(\vec{N}) \quad (1.7)$$

dove $\delta_{i,j}$ è la funzione di Croneker.

- in caso di variabili continue

$$f'(g) = \int d\vec{x} \delta(g - g(\vec{x})) f(\vec{x}) \quad (1.8)$$

dove δ è la funzione di Dirac.

Nel caso di una funzione di una sola variabile è possibile risolvere l'integrale e, facendo uso delle proprietà della funzione di Dirac, calcolare

$$f'(g) = \frac{f(x(g))}{\left. \frac{dg}{dx} \right|_{x=x(g)}} \quad (1.9)$$

1.2 Proprietà della pdf

Non è sempre possibile, ed in generale necessario, conoscere l'intera pdf. Molte informazioni possono essere estratte da funzionali della pdf stessa, descritti nel resto di questa sezione.

1.2.1 moda

La *moda* della distribuzione è il valore di x che massimizza la pdf (cioè il valore che si presenta il maggior numero di volte). Se c'è solo una moda la funzione è unimodulare.

1.2.2 mediana

La *mediana* della pdf è definita come il valore di x per cui

$$\int_{-\infty}^{x_{mediana}} f(x) dx = \frac{1}{2}$$

e cioè quel valore che divide la distribuzione dei dati in due parti di area uguale.

1.2.3 Valore di aspettazione

Il valore di aspettazione (*media*) è una misura del valore al centro della distribuzione.

$$E[x] = \mu = \int x f(x) dx \quad (1.10)$$

proprietà:

- In caso di una funzione di una variabile

$$E[g] = \int g(x) f(x) dx \quad (1.11)$$

Sostituendo lo sviluppo in serie di $g(x)$ nell'intorno di μ (con una sola variabile casuale, per semplicità)

$$E[g] = \int \sum_{k=1, \infty} g^{(k)}(\mu) \frac{(x-\mu)^k}{k!} f(x) dx = g(\mu) + \frac{1}{2} V[x] g^{(2)}(\mu) + \dots \quad (1.12)$$

dove $g^{(k)}$ è la derivata k-sima di g e $V[x]$ è la varianza (vedi paragrafo successivo).

Si vede dunque che, a meno che la varianza di x non è nulla (cioè x assume un solo valore) il valore di aspettazione di g NON è il valore di g nel valore di aspettazione di x , a meno che g non sia lineare. Si usa però questa approssimazione quando la derivata seconda della funzione è piccola.

- valore di aspettazione di una costante

$$E[a] = a \quad a = \text{costante} \quad (1.13)$$

- valore di aspettazione di una combinazione lineare

$$E[a_1g_1(x) + a_2g_2(x)] = a_1E[g_1(x)] + a_2E[g_2(x)] \quad (1.14)$$

1.2.4 Varianza

La misura della dispersione della VC o della funzione $g(x)$ intorno al suo valore di aspettazione è chiamata *varianza*

$$V[x] = E[(x - E[x])^2] = \sigma^2 = \int (x - \mu)^2 f(x) dx \quad (1.15)$$

con σ non negativa, detta *deviazione standard*.

proprietà:

-

$$V[g(x)] = E[(g(x) - E[g(x)])^2] \quad (1.16)$$

Utilizzando la proprietà di linearità del valore aspettato si ricava

$$V[x] = E[(x - \mu)^2] = E[x^2] - (E[x])^2 \quad (1.17)$$

Dimostrazione:

$$V[x] = E(x^2 + \mu^2 - 2x\mu) = E[x^2 + (E[x])^2 - 2xE[x]] \quad (1.18)$$

Utilizzando $E[\sum_i a_i x] = \sum_i a_i E[x]$

$$\begin{aligned} V[x] &= E[x^2] + E[E^2[x]] - 2E[xE[x]] \\ &= E[x^2] + E^2[x] - 2E[x]E[x] \\ &= E[x^2] - E^2[x] \end{aligned} \quad (1.19)$$

- Varianza di combinazioni lineari

$$V(ax) = a^2V(x) \quad a = \text{costante} \quad (1.20)$$

Dimostrazione:

$$V[ax] = E[a^2x^2] - E^2[ax] = a^2E[x^2] - (aE[x])^2 = a^2V[x] \quad (1.21)$$

- Nel caso di più variabili, si definisce varianza della singola variabile la sua variazione a prescindere dal valore delle altre:

$$\sigma_i^2 = E[(x_i - E[x_i])^2] \quad (1.22)$$

Per tenere in conto anche le correlazioni tra le variabili si introduce la *matrice di covarianza*

$$V_{ij} = E[(x_i - E[x_i])(x_j - E[x_j])] \quad (1.23)$$

- $\widehat{V}(\vec{x})$ è simmetrica
- un elemento diagonale V_{ii} è chiamato *varianza* σ_i^2 della variabile x_i

$$\sigma_i^2 > 0 \quad \sigma_i^2 = V_{ii} = E[x_i - E(x_i)]^2 = E(x_i^2) - [E(x_i)]^2$$

- un elemento fuori dalla diagonale, V_{ij} con $i \neq j$, è chiamato *covarianza* di x_i e x_j ed è denotato con $cov(x_i, x_j)$

$$cov(x_i, x_j) = V_{ij} = E(x_i x_j) - E(x_i)E(x_j)$$

Dimostrazione:

$$\begin{aligned} V_{ij} &= E[(x_i - E(x_i))(x_j - E(x_j))] & (1.24) \\ &= E[x_i x_j - E(x_j)x_i - E(x_i)x_j + E(x_i)E(x_j)] \\ &= E(x_i, x_j) - 2E(x_i)E(x_j) + E(x_i)E(x_j) \\ &= E(x_i x_j) - E(x_i)E(x_j) \end{aligned}$$

dunque una variabile influenza l'altra, visto che il valor medio del prodotto è diverso dal prodotto dei valori medi.

Si vede che se $x_i = x_j \Rightarrow V_{ii} = E(x_i^2) - (E(x_i))^2$.

La covarianza può essere una quantità positiva o negativa.

- Una quantità utilizzata frequentemente per la correlazione fra due variabili è il *coefficiente di correlazione* $\rho(x_i, x_j)$

$$\rho(x_i, x_j) = \rho_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}} = \frac{V_{ij}}{\sigma_i \sigma_j} = \frac{cov(x_i, x_j)}{\sigma_i \sigma_j}$$

che soddisfa

$$-1 \leq \rho(x_i, x_j) \leq +1$$

Per comprendere il significato del coefficiente di correlazione si consideri il caso di due variabili (x_1, x_2) e si scriva la *pdf* come

$$f(x_1, x_2) = \Phi_1(x_1)\Phi_2(x_2) + \Delta(x_1, x_2) \quad (1.25)$$

dove si è definito $\Phi_1(x_1) = \int dx_2 f(x_1, x_2)$ e $\Phi_2(x_2) = \int dx_1 f(x_1, x_2)$ e dove si ha, per definizione, $\int f(x_1, x_2) dx_1 dx_2 = 1$. Data questa definizione si ha (a voi la dimostrazione)

$$\int \Delta(x_1, x_2) dx_1 dx_2 = 0 \quad (1.26)$$

e

$$\int \Delta(x_1, x_2) x_i dx_1 dx_2 = 0 \quad (i = 1, 2) \quad (1.27)$$

Il coefficiente di correlazione si può dunque scrivere:

$$\begin{aligned} \rho &= \frac{1}{\sigma_1 \sigma_2} \int (x_1 - E[x_1])\Phi_1(x_1) dx_1 \int (x_2 - E[x_2])\Phi_2(x_2) dx_2 & (1.28) \\ &+ \frac{1}{\sigma_1 \sigma_2} \int (x_1 - E[x_1])(x_2 - E[x_2])\Delta(x_1, x_2) dx_1 dx_2 \\ &= \frac{1}{\sigma_1 \sigma_2} \int x_1 x_2 \Delta(x_1, x_2) dx_1 dx_2 \end{aligned}$$

ed è dunque un indicatore della grandezza della funzione Δ relativamente alla varianza delle variabili.

Nel caso in cui le variabili siano scorrelate

$$f(x_1, x_2) = f_1(x_1)f_2(x_2) = \Phi_1(x_1)\Phi_2(x_2) \quad (1.29)$$

per cui $\Delta = 0$. Il coefficiente di correlazione e' perciò un indice di quanto la *pdf* soddisfi la condizione di scorrelazione 1.29.

- Nel caso poi si voglia calcolare la varianza di una funzione di variabili casuali, sviluppando nell'equazione 1.16 g in serie attorno al valore di aspettazione di x (\hat{x}) e fermandosi al primo ordine in $V[x]/\hat{x}$, si ottiene

$$V[g] = E[g^2] - (E[g])^2 = \sum_{i,j} \frac{dg}{dx_i}(\hat{x}) \frac{dg}{dx_j}(\hat{x}) V_{ij}[\hat{x}] + o((V[\hat{x}]/\hat{x})^2) \quad (1.30)$$

Dove si è ritrovato che, per variabili x con piccola varianza vale la propagazione degli errori comunemente usata. Essa però non è esatta, la cosa più corretta è ottenere l'intera *pdf* di g , come discusso in sezione 1.1.2.

1.3 Funzione di distribuzione cumulativa

Invece di caratterizzare la VC con la sua *pdf* $f(x)$, si può utilizzare la *distribuzione cumulativa* $F(x)$, definita da

$$F(x) = \int_{x_{min}}^x f(x') dx' \quad (1.31)$$

dove x_{min} è il valore limite più piccolo di x .

Dato che $f(x)$ è sempre non negativa, $F(x)$ è chiaramente una funzione monotona crescente di x nell'intervallo $x_{min} \leq x \leq x_{max}$.

La *cumulativa* rappresenta la probabilità che la variabile X , dotata di *pdf* $f(x')$, assuma un valore compreso tra x_{min} e x . Si ha che

$$f(x) = \frac{dF(x)}{dx}$$

Applicando l'equazione 1.31 si ottiene che la *pdf* della cumulativa è

$$f'(F) = \frac{f(x)}{\frac{dF}{dx}} = \frac{f(x)}{f(x)} = 1 \quad (1.32)$$

La cumulativa è distribuita in modo uniforme tra 0 e 1.

1.4 Estimatori

Il risultato di un esperimento è un insieme di variabili casuali. I modelli teorici forniscono delle possibili *pdf* per queste variabili, funzioni di variabili incognite $f(x, \vec{\mu})$.

L'analisi statistica dei dati sperimentali si ripropone di costruire funzioni delle variabili casuali $\mu_i = \mu_i(\vec{x})$ il cui valore di aspettazione sia quanto più possibile prossimo al valore vero dei parametri (μ_i^T). Fattori di merito di un estimatore sono:

- bias è la distanza tra il valore di aspettazione dell'estimatore ed il parametro che si vuole stimare. Esso rappresenta perciò la deviazione dal valore che si vuole misurare dopo infinite misure
- accuratezza è la varianza dell'estimatore e rappresenta la precisione con cui si effettua la misura.

1.5 Campioni di dati con stessa pdf

Il resto di questo lavoro si ripropone di descrivere tecniche per estrarre i parametri delle *pdf* che descrivono i dati. Tali tecniche richiedono in generale che si facciano ipotesi sulla *pdf*. C'è un caso particolare in cui si possono estrarre stime anche in assenza di ipotesi sulla *pdf*: il caso in cui si siano effettuate N misure x_i della stessa variabile casuale x . In questo caso un buon estimatore di $\mu = E[x]$ è

$$\hat{\mu} = \frac{\sum_{i=1}^N x_i}{N} \quad (1.33)$$

dal momento che

$$E[\hat{\mu}] = \frac{\sum_{i=1}^N E[x_i]}{N} = \mu \quad (1.34)$$

e la varianza

$$V[\hat{\mu}] = \frac{\sum_{i=1}^N v[x_i]}{N^2} = \frac{\sigma^2}{N}, \quad (1.35)$$

dove σ^2 è la varianza di x , cala con il numero di eventi.

È inoltre possibile stimare σ . Un buon estimatore è infatti

$$\hat{s}^2 = \frac{\sum_{i=1}^N (x_i - \hat{\mu})^2}{N - 1} \quad (1.36)$$

per il quale

$$E[\hat{s}^2] = \frac{\sum_{i=1}^N E[(x_i - \hat{\mu})^2]}{N - 1} = \frac{NE[x_i^2] - \frac{1}{N}E[\sum_{i,j} x_i x_j]}{N - 1} \quad (1.37)$$

Poichè

$$E[\sum_{i,j} x_i x_j] = N(\sigma^2 - \mu^2) + N(N - 1)\mu = N\sigma^2 + N^2\mu^2 \quad (1.38)$$

$$E[\hat{s}^2] = \frac{N(\sigma^2 - \mu^2) - \sigma^2 + N\mu^2}{N - 1} = \sigma^2 \quad (1.39)$$

e

$$V[\hat{s}^2] = \frac{2\sigma^4}{N - 1} \sim 4\sigma^2 V[\hat{s}] \quad (1.40)$$

Da qui l'espressione approssimata

$$\sigma_s^{\hat{}} = \frac{\sigma^2}{2(N - 1)} \quad (1.41)$$

Dato dunque un campione di eventi con la stessa *pdf* si stimano valore medio e deviazione standard come $\hat{\mu} \pm \frac{\hat{s}}{\sqrt{N}}$ e $\hat{s} \pm \frac{\hat{s}}{2(N-1)}$

file=hist.eps,height=8.cm

1.6 Esercizi

Esercizio 1

Sia SLD che LEP hanno effettuato misure di $x = \sin^2\theta_w$ (v. figura). Trattiamo la misura di SLD come un'unica misura ($x=0.23055\pm 0.00041$) mentre assumiamo che LEP abbia fatto 5 misure con ugual precisione, ignota, ma trascurabile ai fini di questo esercizio:

$x=0.23102, 0.23228, 0.23243, 0.2314, 0.2322$. Valutare se i due esperimenti sono consistenti entro 1σ . Cosa si impara da questa affermazione?

Esercizio 2 Supponiamo di effettuare N misure di una variabile casuale discreta e di ottenerla distribuzione mostrata in figura Stimare valor medio e varianza di questa variabile ed i loro relativi errori.

Capitolo 2

Funzioni di distribuzione

Si discutono in questo capitolo le proprietà delle funzioni di distribuzione più frequenti in Fisica delle Particelle Elementari: funzioni di distribuzione discrete, quali la binomiale e la poissoniana, e funzioni di distribuzione continue, quali la gaussiana e la distribuzione uniforme.

2.1 Distribuzione binomiale

Si applica ogni volta che si effettua un numero N di esperimenti in cui esistono solo due possibilità: successo o fallimento.

Definendo p ($0 \leq p \leq 1$) la probabilità di avere un successo, e $q = 1 - p$ la probabilità di avere un fallimento, in una sequenza di N prove indipendenti la probabilità di avere n successi e $N-n$ fallimenti è

$$B(n; N, p) = \binom{N}{n} p^n (1-p)^{N-n} = B(N-n; N, 1-p) \quad (2.1)$$

Il coefficiente binomiale $\binom{N}{n}$ tiene conto del fatto che non è possibile stabilire l'ordine dei successi e dei fallimenti ed è definito come

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \binom{N}{N-n} \quad (2.2)$$

Si vede facilmente che le probabilità sono normalizzate ad 1:

$$\sum_{n=0}^N B[n; N, p] = \sum_{n=0}^N \binom{N}{n} p^n q^{N-n} = (p+q)^N = 1 \quad (2.3)$$

La distribuzione binomiale è simmetrica quando $p = q = \frac{1}{2}$

$$B(n; N, \frac{1}{2}) = \binom{N}{n} \frac{1}{2^N} \quad (2.4)$$

Valor medio

Il valor medio di una variabile n , distribuita in modo binomiale è

$$\begin{aligned}
 E[n] &= \sum_{n=0}^N nB(n; N, p) & (2.5) \\
 &= \sum_{n=1}^N n \binom{N}{n} p^n (1-p)^{N-n} \\
 &= Np \sum_{n=1}^N n \frac{N-1!}{n!N-n!} p^{n-1} (1-p)^{N-n}
 \end{aligned}$$

Se si ridefinisce $n' = n - 1$ ed $N' = N - 1$

$$E[n] = Np \sum_{n'=0}^{N'} \frac{N'!}{n'!N'-n'!} p^{n'} (1-p)^{N'-n'} \quad (2.6)$$

e visto che la sommatoria è proprio la normalizzazione di una binomiale con probabilità p ed N' elementi,

$$E[n] = Np \quad (2.7)$$

Varianza

Per calcolare la varianza di una variabile n distribuita in modo binomiale si utilizza l'equazione 1.17. Si deve perciò calcolare

$$\begin{aligned}
 E[n^2] &= \sum_{n=0}^N n^2 B[n; N, p] & (2.8) \\
 &= \sum_{n=1}^N n^2 \binom{N}{n} p^n (1-p)^{N-n} \\
 &= Np \sum_{n=1}^N n^2 \frac{N-1!}{n!N-n!} p^{n-1} (1-p)^{N-n}
 \end{aligned}$$

Se si ridefinisce $n' = n - 1$ ed $N' = N - 1$

$$\begin{aligned}
 E[n^2] &= Np \sum_{n'=0}^{N'} (n'+1) \frac{N'!}{n'!N'-n'!} p^{n'} (1-p)^{N'-n'} & (2.9) \\
 &= Np(N'p + 1) = Np(Np - p + 1)
 \end{aligned}$$

dove si è utilizzata sia l'equazione 2.7, che la definizione di normalizzazione.

Si ha dunque

$$V[x] = E[n^2] - (E[n])^2 = Np(Np - p + 1) - (Np)^2 = Np(1-p) = Npq \quad (2.10)$$

Spesso si è interessati alla quantità $\frac{n}{N}$, il numero relativo di successi in N prove. In questo caso

$$E\left[\frac{n}{N}\right] = \frac{1}{N} E[n] = p \quad (2.11)$$

$$V\left[\frac{n}{N}\right] = \left(\frac{1}{N}\right)^2 V[n] = \frac{p(1-p)}{N} = \frac{pq}{N} \quad (2.12)$$

Il parametro p ha dunque il significato di probabilità di successo, visto che è il valore di aspettazione (ad infiniti tentativi) della frazione di successi, $\frac{n}{N}$. Nel limite di infiniti tentativi ($N \rightarrow \infty$) la varianza di questa frazione è nulla.

2.1.1 Applicazione: l'efficienza di selezione

Supponiamo di voler discriminare tra un segnale ed i suoi fondi tramite una variabile x , richiedendo che si consideri un evento se è soddisfatta la selezione $x > X_{cut}$. Si definisce efficienza ϵ della selezione la frazione di eventi di segnale, nel limite di infiniti eventi prodotti, che soddisfano la selezione stessa.

Poichè un evento di segnale o passa la selezione oppure non la passa, il numero di eventi di segnale selezionati, N_{sel} , è distribuito in modo binomiale. Il parametro p , che abbiamo visto rappresentare la frazione di successi dopo infiniti tentativi, è per definizione l'efficienza.

$$P(N_{sel}|N_{tot}, \epsilon) = B(n; N_{tot}, \epsilon) \quad (2.13)$$

Si ha dunque che un buon estimatore dell'efficienza è $\hat{\epsilon} = \frac{N_{sel}}{N_{tot}}$, in quanto $E[\hat{\epsilon}] = \epsilon$. L'errore su questo estimatore è

$$\sigma_{\hat{\epsilon}}^2 = V[\hat{\epsilon}] = V\left[\frac{N_{sel}}{N_{tot}}\right] = \frac{V[N_{sel}]}{N_{tot}^2} = \frac{N_{tot}\epsilon(1-\epsilon)}{N_{tot}^2} \quad (2.14)$$

e dunque

$$\sigma_{\hat{\epsilon}} = \sqrt{\frac{\epsilon(1-\epsilon)}{N_{tot}}} \quad (2.15)$$

Solitamente si usano eventi simulati (MonteCarlo, vedi capitolo 6) per stimare le efficienze come rapporto tra gli eventi selezionati ed il numero totale di eventi simulati. Si può dunque notare che l'errore sull'efficienza (equazione 2.15) decresce con l'aumentare del numero di eventi simulati.

2.2 Distribuzione di Poisson

La distribuzione di Poisson è il limite della distribuzione binomiale nel caso si abbia un numero molto elevato di tentativi ($N \rightarrow \infty$) ed una probabilità di successo molto piccola ($p \rightarrow 0$) ma con $N \cdot p = \mu = cost$. Il classico esempio è il decadimento radiativo di un materiale. Il numero di atomi che possono decadere (N) è elevato ma la probabilità di decadimento bassa. Nel complesso il numero medio di decadimenti nell'unità di tempo è un numero finito ed osservabile (per materiali sufficientemente radiattivi).

Per effettuare il limite bisogna ricordare che

$$\lim_{N \rightarrow \infty} \left(1 - \frac{\mu}{N}\right)^N = e^{-\mu} \quad (2.16)$$

e che, per la formula di Stirling

$$\lim_{N \rightarrow \infty} N! = \sqrt{2\pi N} N^N e^{-N} \quad (2.17)$$

Si ha perciò

$$\begin{aligned}
 P_\mu[n] &= P(n; \mu) = \lim_{N \rightarrow \infty, p = \frac{\mu}{N}} \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} & (2.18) \\
 &= \lim_{N \rightarrow \infty} \frac{1}{n!} \frac{\sqrt{2\pi N} N^N e^{-N}}{\sqrt{2\pi(N-n)} (N-n)^{N-n} e^{-(N-n)}} \left(\frac{\mu}{N}\right) \left(1 - \frac{\mu}{N}\right)^{N-n} \\
 &= \lim_{N \rightarrow \infty} \frac{1}{n!} \frac{1}{\left(1 - \frac{\mu}{N}\right)^N e^{\mu}} \mu^n \left(1 - \frac{\mu}{N}\right)^N
 \end{aligned}$$

ed in definitiva

$$\Rightarrow P(n; \mu) = \frac{1}{n!} e^{-\mu} \mu^n \quad (2.19)$$

Valor medio

Il valor medio di una variabile n distribuita in modo poissoniano con media μ è

$$E[n] = \sum_{n=0}^{\infty} n P(n; \mu) = \sum_{n=1}^{\infty} n \frac{1}{n!} e^{-\mu} \mu^n = \mu \sum_{n=1}^{\infty} \frac{1}{(n-1)!} e^{-\mu} \mu^{n-1} \quad (2.20)$$

Se si definisce $n' = n - 1$

$$E[n] = \mu \sum_{n'=0}^{\infty} \frac{1}{n'!} e^{-\mu} \mu^{n'} = \mu \quad (2.21)$$

dove si è utilizzata la normalizzazione della distribuzione Poissoniana.

Varianza

Per ottenere la varianza di una variabile n distribuita in modo poissoniano con media μ si utilizza l'equazione 1.17. Si deve perciò calcolare

$$E[n^2] = \mu \sum_{n=1}^{\infty} n \frac{1}{(n-1)!} e^{-\mu} \mu^{n-1} = \mu \sum_{n'=0}^{\infty} (n'+1) \frac{1}{n'!} e^{-\mu} \mu^{n'} = \mu(\mu+1) \quad (2.22)$$

e dunque

$$V[n] = E[n^2] - (E[n])^2 = \mu(\mu+1) - (\mu)^2 = \mu \quad (2.23)$$

Si ha dunque che la varianza è numericamente uguale al valor medio. Si ha perciò che la larghezza relativa della distribuzione, stimata con $\frac{\sqrt{V[n]}}{E[n]} = \frac{1}{\sqrt{\mu}}$ è inversamente proporzionale al valor medio.

Poichè vale l'equazione 2.21, n è un estimatore senza bias di μ e l'errore relativo su di esso è tanto migliore quanto maggiore è n .

La funzione di distribuzione di Poisson è molto asimmetrica per piccoli μ ed ha una coda per n maggiori del valor medio.

2.2.1 Somma di variabili poissoniane

Date due variabili casuali discrete k_1, k_2 con distribuzioni poissoniane indipendenti tra loro di media μ_1, μ_2 rispettivamente, ci si chiede quale sia la distribuzione della variabile $k = k_1 + k_2$.

La distribuzione congiunta delle due variabili è:

$$g(k_1, k_2) = e^{-\mu_1} \frac{\mu_1^{k_1}}{k_1!} e^{-\mu_2} \frac{\mu_2^{k_2}}{k_2!}$$

La distribuzione della somma delle variabili si ottiene sommando su tutte le combinazioni che hanno lo stesso valore di k (vedi paragrafo 1.1.2).

$$\begin{aligned} g(k) &= \sum_{k_1, k_2} \delta(k_1 + k_2 - k) g(k_1, k_2) & (2.24) \\ &= \sum_{k_1, k_2} \delta(k_1 + k_2 - k) e^{-\mu_1} \frac{\mu_1^{k_1}}{k_1!} e^{-\mu_2} \frac{\mu_2^{k_2}}{k_2!} \\ &= e^{-(\mu_1 + \mu_2)} \sum_{k_1} \frac{\mu_1^{k_1}}{k_1!} \frac{\mu_2^{k-k_1}}{k-k_1!} \\ &= e^{-(\mu_1 + \mu_2)} \sum_{k_1} \frac{\mu_1^{k_1}}{k_1!} \frac{(\mu - \mu_1)^{k-k_1}}{k-k_1!} \end{aligned}$$

Introducendo $p = \mu_1/\mu$, $q = 1 - p$ e $\mu = \mu_1 + \mu_2$

$$g(k) = e^{-(\mu_1 + \mu_2)} \sum_{k_1} \mu^k \frac{p^{k_1} q^{k-k_1}}{k_1!(k-k_1)!} = e^{-\mu} \frac{\mu^k}{k!} \sum_{k_1} \frac{k!}{k_1!(k-k_1)!} p^{k_1} q^{k-k_1} = e^{-\mu} \frac{\mu^k}{k!} \quad (2.25)$$

si ottiene che k è distribuita in modo poissoniano con valor medio pari alla somma dei valori medi.

2.2.2 Applicazione: l'istogramma

Un insieme di variabili casuali che si incontra spesso in Fisica delle Particelle Elementari è il cosiddetto istogramma. Date N misure $x_j, j = 1, N$ di una variabile casuale x distribuita con *pdf* $f(x)$, e' possibile suddividere i valori assumibili da x in N_B intervalli (bin $I_i, i = 1, N_B$) e contare quante delle N misure cadono nel bin i -esimo (N_i , con $\sum_{i=1}^{N_B} N_i = N$). Queste N_i sono a loro volta variabili casuali distribuite in modo poissoniano con valore di aspettazione

$$\mu_i = \int_{I_i} f[\xi] d\xi \quad (2.26)$$

per cui

$$P(N_i | \mu_i) = \frac{\mu_i^{N_i} e^{-\mu_i}}{N_i!} \quad (2.27)$$

Questo è vero nell'ipotesi di avere un numero di bin sufficientemente elevato (qualche decina). Bisogna infatti che la probabilità

che un evento cada in un bin piuttosto che in ogni altro sia sufficientemente piccola e che il numero totale di eventi presenti nell'istogramma sia sufficientemente elevato. Altrimenti le N_i sarebbero distribuite secondo la distribuzione multinomiale.

2.3 Distribuzione uniforme

Distribuzione continua più semplice che si può immaginare è la distribuzione uniforme, costante nell'intervallo $[a, b]$ in cui x è definita

$$f(x) = \text{cost} \quad (2.28)$$

Imponendo la normalizzazione

$$\int_x f(x) = 1 \quad (2.29)$$

ai ha

$$\text{cost} = \frac{1}{b-a} \quad (2.30)$$

Valore atteso

Il valor medio di una variabile x distribuita in modo uniforme nell'intervallo $[a, b]$ è

$$E[x] = \int_a^b x f[x] dx = \int_a^b \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{1}{2}(b+a) \quad (2.31)$$

Varianza

La varianza di una variabile x distribuita in modo uniforme nell'intervallo $[a, b]$ è

$$\begin{aligned} V[x] &= \int_a^b (x - E[x])^2 f[x] dx & (2.32) \\ &= \int_a^b \frac{1}{b-a} \left(x - \frac{b+a}{2}\right)^2 dx \\ &= \frac{1}{b-a} \int_a^b \left(x^2 + \frac{(b+a)^2}{4} - x(b+a)\right) dx \\ &= \left(\frac{b+a}{2}\right)^2 + \frac{1}{b-a} \frac{1}{3} (b^3 - a^3) - \frac{1}{b-a} \frac{b+a}{2} (b^2 - a^2) \\ &= \frac{1}{12} (b-a)^2 \end{aligned}$$

Quando $b = 1$ ed $a = 0$ si ha

$$E[x] = \frac{1}{2} \quad (2.33)$$

$$V[x] = \frac{1}{12} \quad (2.34)$$

$$\sigma_x = \frac{1}{\sqrt{12}} \quad (2.35)$$

2.3.1 Applicazione: packing dei dati.

Una tipica applicazione della distribuzione uniforme è il “packing” (impacchettamento) dei dati. Quando si devono immagazzinare i dati, di per sè essi vengono scritti in parole di 2 o 4 byte (cioè parole di 8 o 16 bit, ossia che possono assumere 2^8 o 2^{16} valori differenti). Spesso la risoluzione sulle quantità che si vogliono scrivere è tale che un numero di valori così largo sia eccessivo. In questi casi si suddivide l’intervallo possibile della variabile che si vuole immagazzinare in tanti sotto intervalli (bin) e si salva come informazione quale bin contiene il valore misurato.

Quando si leggono i dati si sa soltanto che originariamente la variabile si trovava in un dato bin. Detti dunque gli estremi del bin in questione x_i e $x_i + W_{bin}$, dove W_{bin} è la larghezza del bin, la variabile x è distribuita in modo uniforme tra questi due estremi. Il valore di aspettazione del suo valore è dunque

$$E[X] = x_i + \frac{W_{bin}}{2} \quad (2.36)$$

e la sua varianza

$$V[X] = \frac{W_{bin}^2}{12} \quad (2.37)$$

Si sceglie dunque la larghezza del bin in modo che l’incertezza aggiuntiva indotta dal packing, $\sigma_{packing} = \frac{W_{bin}}{\sqrt{12}}$, sia trascurabile rispetto alla risoluzione intrinseca della variabile.

2.4 Distribuzione normale o gaussiana

Nel limite di un numero elevato di eventi n , la distribuzione Poissoniana viene approssimata dalla distribuzione Gaussiana.

$$\lim_{n \rightarrow \infty} P_\mu(n) = \lim_{n \rightarrow \infty} \frac{e^{-\mu} \mu^n}{n!} = \lim_{n \rightarrow \infty} \frac{e^{(n-\mu)} (\frac{\mu}{n})^n}{\sqrt{2\pi n}} \quad (2.38)$$

dove si è fatto uso della formula di Stirling 2.17. Se si esprime n in termini del suo valor medio e varianza e si definisce il *pull* $\xi = \frac{n-E[n]}{\sqrt{V[n]}} = \frac{n-\mu}{\sqrt{\mu}}$, sostituendo $\mu \sim n - \xi\sqrt{\mu}$ (sviluppo al primo ordine in $\frac{\xi}{n}$) si ottiene

$$\lim_{n \rightarrow \infty} P_\mu(n) = \lim_{n \rightarrow \infty} \frac{e^{\xi\sqrt{\mu}}}{\sqrt{2\pi\mu}} \left(1 - \frac{\xi}{\sqrt{n}}\right)^n = \lim_{n \rightarrow \infty} \frac{e^{\xi(\sqrt{\mu}-\sqrt{n})}}{\sqrt{2\pi\mu}} \quad (2.39)$$

Sostituendo nuovamente l'espressione di $n = \mu + \xi\sqrt{\mu}$ e sviluppando al secondo ordine in $\frac{\xi}{n}$ si ottiene

$$\lim_{n \rightarrow \infty} P_\mu(n) = \frac{e^{-\frac{\xi}{2}}}{\sqrt{2\pi\mu}} \quad (2.40)$$

Dato che non c'è alcun motivo di limitarsi allo spazio degli interi, sostituendo l'espressione per ξ in termini di $E[x] = \mu$ e $\sqrt{V[X]} = \sigma$, possiamo scrivere la PDF normale come

$$G[x] = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \quad (2.41)$$

2.4.1 Proprietà distribuzione Gaussiana

Per costruzione il valore di aspettazione di una variabile distribuita in modo Gaussiano è $E[x] = \mu$ e la sua varianza $V[x] = \sigma^2$

La *pdf* Gaussiana è simmetrica intorno a $x = \mu$, e quindi la mediana coincide con la media.

La *pdf* gaussiana è tale che la frazione di eventi contenuta in $x \in [\mu - N\sigma, \mu + N\sigma]$ è calcolabile. In particolare per $N = 1$ la frazione è il 32%.

2.4.2 pdf Gaussiana a più variabili

Nel caso di più variabili (N_D), $\vec{x} \in \mathfrak{R}^{N_D}$ si ha

$$G[\vec{x}] = N e^{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \widehat{V}^{-1}(\vec{x} - \vec{\mu})} \quad (2.42)$$

La costante di normalizzazione si trova imponendo

$$\int G[\vec{x}] dx_1, \dots, dx_{N_D} = 1 \quad (2.43)$$

$$N = \frac{1}{(\sqrt{2\pi})^{N_D} \det \widehat{V}} \quad (2.44)$$

In questo i valori medi e la matrice di covarianza (equazione 1.23) sono

$$E[x_i] = \mu_i \quad (2.45)$$

$$V[x_i, x_j] = V_{ij} \quad (2.46)$$

Si definiscono errori delle singole variabili

$$\sigma_i = \sqrt{V_{ii}} \quad (2.47)$$

e la matrice di correlazione diventa

$$\rho_{ij} = \frac{V_{ij}}{\sigma_i \sigma_j} \quad (2.48)$$

Per comprendere il significato di coefficiente di correlazione, consideriamo il caso di due variabili ($N_D = 2$) e supponiamo di aver misurato una delle due variabili e che il risultato disti di $\xi\sigma$ dal valore vero: $x_1 = \mu_1 + \xi\sigma_1$. Esplicitando la matrice di covarianza il valor medio della seconda variabile è

$$\begin{aligned} E[x_2]_{x_1=\mu_1+\xi\sigma_1} &= \mu'_2 = N \int_{-\text{inf}}^{+\text{inf}} G'(x_2) x_2 dx_2 \quad (2.49) \\ &= \int_{-\text{inf}}^{+\text{inf}} N e^{-\frac{1}{2(1-\rho_{12}^2)} \left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - 2\rho_{12} \frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right)} x_2 dx_2 \end{aligned}$$

dove la normalizzazione N si ottiene imponendo

$$\int_{-\text{inf}}^{+\text{inf}} N G'(x_2) dx_2 = 1 \quad (2.50)$$

Ma poichè, avendo definito $\xi_2 = \frac{x_2-\mu_2}{\sigma_2}$

$$\begin{aligned} \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - 2\rho_{12} \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} \quad (2.51) \\ = \xi^2 + \xi_2^2 - 2\rho_{12}\xi\xi_2 = (\xi_2 - \rho_{12}\xi)^2 + \xi^2(1 - \rho_{12}^2) \end{aligned}$$

si può scrivere

$$G(x_2) = e^{-\frac{\xi^2}{2}} e^{-\frac{(\xi_2 - \rho_{12}\xi)^2}{2(1-\rho_{12}^2)}} = e^{-\frac{\xi^2}{2}} e^{-\frac{(x_2 - \mu_2 - \rho_{12}\xi\sigma_2)^2}{2(1-\rho_{12}^2)\sigma_2^2}} \quad (2.52)$$

G è dunque proporzionale ad una gaussiana con valor medio $\mu_2 + \rho_{12}\xi\sigma_2$ per cui

$$N = \frac{e^{-\frac{\xi^2}{2}}}{\sqrt{2\pi\sigma_2}} \quad (2.53)$$

e

$$\mu'_2 = \mu_2 + \rho_{12}\xi\sigma_2 \quad (2.54)$$

Perciò, se si è misurato per uno delle due variabili un valore che è spostato rispetto a quello vero di ξ sigma, l'altra variabile che ha con essa una correlazione ρ_{12} sarà in media spostata rispetto al suo valore centrale di $\rho_{12}\xi$ sigma.

2.5 Teorema del limite centrale

Considerando N variabili casuali x_i , $i = 1, \dots, N$ indipendenti, ciascuno con il suo valor medio μ_i e la sua varianza σ_i^2 , di cui non si conosce però la pdf $f(x_i)$, nel limite per $N \rightarrow \infty$ la pdf della variabile $x = \frac{\sum_i x_i}{N}$ è tale che

$$\lim_{N \rightarrow \infty} f(x) = G_{\mu, \sigma}[x] \quad (2.55)$$

Si ha cioè una pdf gaussiana con

$$\mu = \frac{\sum_i \mu_i}{N} \quad (2.56)$$

$$\sigma = \sqrt{V[x]} = \sqrt{\frac{\sum_i V[x_i]}{N^2}} = \sqrt{\frac{\sum_i \sigma_i^2}{N^2}} \quad (2.57)$$

Questo teorema viene utilizzato per esempio in Fisica delle Particelle Elementari per definire l'errore sistematico: ogni esperimento fisico include molte sorgenti di errori e l'osservabile che si misura può essere considerato come la somma di un "osservabile vero", cioè la quantità di cui si vuole misurare il valore vero, e un "errore totale", proveniente dalla combinazione di un numero ignoto di errori elementari indipendenti.

Quando il numero di sorgenti di errore diventa grande, l'errore totale si assume avere una distribuzione gaussiana:

$$x \longrightarrow x_T + \sum_i \varepsilon_i \longrightarrow x_T + \xi \quad (2.58)$$

$$f(\xi) = G_{\mu, \sigma}[\xi] \quad (2.59)$$

dove μ è lo spostamento rispetto al valore aspettato (bias)

2.6 Esercizi

Esercizio 3

Probabilità a priori della pesca di beneficenza. Supponiamo che siano stati venduti N biglietti e che io ne abbia acquistati n . Calcolare

- A) la probabilità di vittoria dopo una singola estrazione
 B) la probabilità di vincere con almeno un biglietto dopo due estrazioni Discutere i casi notevoli in cui io abbia comprato tutti i biglietti ($n = N$) o quello in cui si sia comprato un solo biglietto ($n = 1$)

Esercizio 4

Supponiamo che in un evento siano stati prodotti quattro quark b . Supponiamo inoltre di essere in grado di identificare leptoni (l) e D^* . Assumendo che il branching ratio semileptonico sia $\text{Br}(B \rightarrow l) = 20\% = B_{SL}$, mentre $\text{Br}(B \rightarrow D^*l) = 25\% = B_{D^*}$ e $\text{Br}(B \rightarrow D^*l\nu) = 5\% = B_{D^*l}$, calcolare la probabilità di avere in un evento:

- A) almeno un leptone ed esattamente un leptone
 B) almeno due leptoni
 C) quale sia il numero di D^* e leptoni che è più probabile trovare in evento, e determinare detta probabilità.

Esercizio 5 Date due variabili casuali discrete k_1, k_2 con distribuzioni poissoniane indipendenti tra loro di media μ_1, μ_2 rispettivamente, trovare la distribuzione della variabile $k = k_1 + k_2$.

Esercizio 6 -Poisson e binomiale

A)

Trovare la probabilità che essendo il valore atteso (Poissoniano) di un osservabile μ si osservi N . Calcolare il valore numerico per il caso particolare ($\mu = 3.7; N = 1$).

B)

Supponendo che μ sia stato determinato con un altro campione (per esempio una simulazione), e che in esso si siano trovati 370 eventi su 1000 ($\mu = \epsilon k, k = 10, \epsilon = \frac{N_{sel}}{N_{obs}}$) Quale è la probabilità di aver misurato 1 ?

Esercizio 7 - Misura B.R. $B \rightarrow J\Psi K_s$

Dopo aver raccolto una luminosità integrata $L=10 \text{ fb}^{-1}$ l'analisi di eventi $B \rightarrow J\Psi K_s (J\Psi \rightarrow ll, K_s \rightarrow \pi\pi)$ seleziona 100 candidati. Usando 1000 eventi simulati valuto che l'efficienza è il 37%. Qual è il B.R. misurato, quale l'errore statistico e quale l'errore sistematico derivante dal fatto di aver simulato troppi pochi eventi. ($\sigma_{bb} = 1 \text{ nb}^{-1}$)

Esercizio 8

Un decadimento di un bosone Z^0 in due quark (di carica $\pm q$) si manifesta come due getti di particelle emessi in direzioni opposte. E' possibile definire una variabile (x) per ogni jet che è distribuita con media $\mu (> 0)$ e sigma σ , se il quark generante è quello positivo, media $-\mu$ e stessa risoluzione σ altrimenti.

A) Se uno dei jet ha $x > 0$, trovare la probabilità che il quark generante sia quello positivo.

B) In un evento è certo che se un jet viene da un quark positivo l'altro viene da un quark negativo. Se si è misurato che il jet che è andato in avanti ha X_F maggiore dell'altro ($X_F > X_B$), trovare la probabilità che il quark in avanti sia positivo

C) utilizzando un campione di eventi con coppie di jet (X_F^i, X_B^i) , determinare il valore di μ e σ . Si assuma di saper stimare la correlazione (ρ_{+-}) tra le misure nei due jet.

Capitolo 3

Statistica di Bayes

3.1 Teorema di Bayes

Si consideri di aver effettuato un'osservazione O e che ci siano N possibili ipotesi su di essa, e che questo set di ipotesi sia esauriente:

$$\sum_{i=1}^N P(H_i) = 1 \quad (3.1)$$

Sono note le probabilità che data ogni ipotesi, l'osservazione O avvenga. Queste probabilità, indicate con $P(O|H_i)$ sono dette probabilità a priori. Le *pdf* fin qui discusse sono effettivamente probabilità a priori. Per esempio la distribuzione gaussiana di una variabile x indica la probabilità a priori che la variabile assuma un dato valore sotto l'ipotesi che i parametri (μ e σ) abbiano dati valori. In termini bayesiani

$$g(x; \mu, \sigma) = P(x|\mu, \sigma) \quad (3.2)$$

Affettuare una misura vuol dire stimare valor medio e varianza di un osservabile. Questo non è però esaustivo perchè in questo modo si estrae la probabilità che l'osservabile abbia un dato valore dato il suo valor vero, ma NON la probabilità che il valore vero abbia un determinato valore data l'osservazione. Il teorema di Bayes si ripropone di estrarre la cosiddetta probabilità a posteriori che il valore vero di un osservabile assuma un dato valore, a partire dalla misura fatta.

In termini di osservazioni e probabilità la *Probabilità a posteriori* è indicata con $P(H_i|O)$ e rappresenta la probabilità che sia vera l'ipotesi H_i data l'osservazione O .

Il teorema di Bayes [?] dice che

$$P(H_i|O) = \frac{P(O|H_i)P(H_i)}{\sum_{j=1}^n P(O|H_j)P(H_j)} R \quad (3.3)$$

Dimostrazione:

ricordando la definizione di probabilità composta

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad (3.4)$$

che in termini di ipotesi ed osservazioni è

$$P(H_i|O)P(O) = P(O|H_i)P(H_i) \quad (3.5)$$

e visto che $\sum_{i=1}^n P(H_i|O) = 1$,

$$P(H_i|O) = \frac{P(O|H_i)P(H_i)}{P(O)} = \frac{P(O|H_i)P(H_i)}{\sum_{j=1}^n P(O|H_j)P(H_j)} \quad (3.6)$$

Esempio:

Si considerino tre insiemi A, B e C contenenti 2 monete; A ha 2 monete d'oro, B ha una moneta d'oro e una d'argento, C 2 monete d'argento.

Si scelga un insieme a caso e si prenda una moneta da questo. Supponendo che la prima moneta estratta sia d'oro, qual'è la probabilità che la seconda moneta estratta dallo stesso insieme sia anch'essa d'oro?

Si definisca ipotesi H_1 il fatto di aver scelto l'insieme A e H_2 o H_3 il fatto di aver scelto B o C rispettivamente.

Se O denota l'evento della prima estrazione, noi vogliamo calcolare $P(H_1|O)$. La probabilità a priori $P(O|H_i)$ di prendere una moneta d'oro da H_i è

$$P(O|H_1) = 1 \quad P(O|H_2) = \frac{1}{2} \quad P(O|H_3) = 0 \quad (3.7)$$

inoltre, dato che l'insieme è selezionato a caso,

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3} \quad (3.8)$$

Dal teorema di Bayes si ricava

$$P(H_1|O) = \frac{P(O|H_1)P(H_1)}{\sum_{j=1}^3 P(O|H_j)P(H_j)} = \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3} \quad (3.9)$$

Così, sebbene la probabilità di selezionare H_1 era solo $\frac{1}{3}$, l'osservazione che la prima moneta estratta è d'oro effettivamente raddoppia la probabilità che sia selezionato l'insieme H_1 .

3.1.1 Applicazione alle misure

Il teorema di Bayes può essere applicato al concetto di misura per trovare la probabilità che il valore vero assuma un dato valore, data un'osservabile. Considerando il caso di una distribuzione gaussiana di una variabile x di valore vero x_T . La distribuzione a priori di x_T è, in base alla equazione 3.2

$$P(x|\mu, \sigma) = g(x; \mu\sigma) \quad (3.10)$$

Applicando il Teorema di Bayes è possibile estrarre la probabilità che μ abbia un dato valore una volta effettuata la misura x :

$$P(\mu|x, \sigma) = \frac{P(x|\mu, \sigma)P(\mu)}{\int d\mu P(x|\mu, \sigma)P(\mu)} \quad (3.11)$$

Non abbiamo idea di cosa sia $P(\mu)$: il Postulato di Bayes dice che se la distribuzione della probabilità a priori è completamente sconosciuta, la cosa ottimale da fare è porre $P(\mu) = 1$.

3.1.2 Esempio: misura di vita media

Consideriamo il decadimento di una particella instabile. La *pdf* del tempo t che intercorre tra la produzione ed il decadimento della particella è, in termini della vita media τ :

$$f(t) = \Gamma e^{-\Gamma t} \quad \Gamma = \frac{1}{\tau} \quad (3.12)$$

Se si eseguono N misure di tempo t_i e si definisce tempo medio

$$\bar{t} = \frac{\sum_i t_i}{N}$$

, il suo valore di aspettazione è

$$E(\bar{t}) = \frac{\sum_i E(t_i)}{N} = \frac{1}{\Gamma}$$

con

$$E(t_i) = \Gamma \int e^{-\Gamma t} t dt = \frac{1}{\Gamma} \quad (3.13)$$

e dunque un buon estimatore di Γ è

$$\hat{\Gamma} = \frac{1}{\frac{\sum_i t_i}{N}} \quad (3.14)$$

La varianza di $\frac{1}{\hat{\Gamma}}$ è

$$V\left[\frac{1}{\hat{\Gamma}}\right] = \sum_i \frac{V[t_i]}{N} = \frac{1}{N} \left(\int t^2 \Gamma e^{-\Gamma t} dt - \frac{1}{\hat{\Gamma}^2} \right) = \frac{1}{\hat{\Gamma}^2} \quad (3.15)$$

Dal punto di vista frequentistico nulla si può dire sulla distribuzione di Γ . Si può al più supporre che abbia la stessa risoluzione del suo estimatore:

$$V[\Gamma] \sim V[\hat{\Gamma}] = V\left[\frac{1}{\hat{\Gamma}}\right]\Gamma^4 = \frac{\Gamma^2}{N} \quad (3.16)$$

Nell'approccio Bayesiano, invece, si considerano le misure una alla volta. Dopo la misura del primo tempo t_1 la sua probabilità a priori è

$$f(t) = \Gamma e^{-\Gamma t} = P(t|\Gamma)_1 \quad (3.17)$$

Applicando il Postulato di Bayes, la *pdf* di Γ è

$$P(\Gamma|t_1) = \frac{P(t_1|\Gamma)P(\Gamma)}{\int d\Gamma P(t_1|\Gamma)P(\Gamma)} = \frac{\Gamma e^{-\Gamma t_1}}{\int d\Gamma \Gamma e^{-\Gamma t_1}} = \Gamma e^{-\Gamma t_1} t_1^2 \quad (3.18)$$

Il massimo di questa *pdf*, ossia il valore più probabile per Γ è $\Gamma_{max} = \frac{1}{t_1}$ cioè $\Gamma_{max} = \hat{\Gamma}$. Il valor medio è però

$$E[\Gamma] = \int \Gamma P(\Gamma|t_1) d\Gamma = \int \Gamma^2 t_1^2 e^{-\Gamma t_1} \quad (3.19)$$

Facendo uso della relazione, molto usata nel seguito,

$$\int x^m e^{-ax} dx = \frac{m!}{a^{m+1}} \quad (3.20)$$

si ottiene

$$E[\Gamma] = \frac{2}{t_1} \neq \hat{\Gamma} \quad (3.21)$$

Analogamente si può calcolare la varianza

$$V[\Gamma] = \int \Gamma^2 P(\Gamma|t_1) d\Gamma = \frac{2}{t_1^2} \quad (3.22)$$

differente dalla stima frequentistica in equazione 3.16.

Si consideri ora una seconda misura, t_2 , e si applichi nuovamente il teorema di Bayes:

$$P(\Gamma|t_1, t_2) = \frac{P(t_2|\Gamma)P(\Gamma)}{\int d\Gamma P(t_2|\Gamma)P(\Gamma)} \quad (3.23)$$

e, prendendo come probabilità a priori per Γ la probabilità a posteriori dopo la prima misura:

$$= \frac{P(t_2|\Gamma)P(\Gamma|t_1)}{\int d\Gamma P(t_2|\Gamma)P(\Gamma|t_1)} = \frac{\Gamma e^{-\Gamma t_2} \Gamma t_1^2 e^{-\Gamma t_1}}{\int d\Gamma \Gamma e^{-\Gamma t_2} \Gamma t_1^2 e^{-\Gamma t_1}} = \frac{1}{2} \Gamma^2 (t_1 + t_2)^3 e^{-\Gamma(t_1 + t_2)} \quad (3.24)$$

Il massimo dopo due misure è

$$\Gamma_{max} = \frac{t_1 + t_2}{2} = \hat{\Gamma} \quad (3.25)$$

nuovamente uguale al caso frequentistico, ma valor medio e varianza sono

$$E[\Gamma] = \frac{t_1 + t_2}{3} \quad (3.26)$$

e

$$V[\Gamma] = \frac{(t_1 + t_2)^2}{3} \quad (3.27)$$

diversi dal caso frequentistico.

Nel caso generale, con N misure,

$$P(\Gamma|t_1, \dots, t_N) = \frac{1}{N+1!} \Gamma^N \left(\sum_{k=1}^N t_k \right) (N+1) e^{-\Gamma \sum_{k=1}^N t_k} \quad (3.28)$$

e dunque:

$$\Gamma_{max} = \frac{\sum_{k=1}^N t_k}{N} = \hat{\Gamma} \quad (3.29)$$

nuovamente uguale al caso frequentistico, ma valor medio e varianza sono

$$E[\Gamma] = \frac{\sum_{k=1}^N t_k}{N+1} \quad (3.30)$$

e

$$V[\Gamma] = \frac{(\sum_{k=1}^N t_k)^2}{N+1} \quad (3.31)$$

uguali al caso frequentistico per N grandi.

3.2 Esercizi

Esercizio 9

Dati N rivelatori di particelle in grado di distinguere tra K diversi tipi di particelle ($P(x_i|k)(i = 1, N; k = 1, K)$ e x_i sia una singola misura) :

A)

determinare, nell'ipotesi che tutti i rivelatori abbiano dato una risposta senza correlazioni, la probabilità a posteriori di essere una data particella

B) Discutere il caso specifico in cui si abbiano due rivelatori. Nel primo rivelatore un π da un segnale con $\mu = 0$ e $\sigma = 1$ mentre un K da un segnale con $\mu = \mu_1 = 2$ e $\sigma = 1$. Nel secondo rivelatore il segnale del π è uguale ($\mu = 0, \sigma = 1$) ma il segnale del K è più

separato ($\mu = \mu_2 = 10, \sigma = 1$). determinare la probabilità che una particella sia un pione se il rivelatore 1 ha effettuato una misura ξ_1 tra $(0, \mu_1)$ e che il rivelatore 2 ha misurato $0 < \xi_2 < \mu_2$.

Capitolo 4

Tecniche di fit

4.1 Verosimiglianza (*likelihood*)

Si definisce verosimiglianza (*likelihood*) di ottenere una misura x dato un parametro θ la probabilità di aver misurato x se θ è il valore corretto per quel parametro. Essa è dunque numericamente equivalente alla *pdf*, anche se dal punto di vista interpretativo la *pdf* è la distribuzione vera sottostante, mentre la verosimiglianza è la forma che si ipotizza per essa.

Vedremo in questo capitolo come dalla conoscenza della verosimiglianza per un insieme di dati è possibile estrarre estimatori per i parametri. A tal fine è necessario che la *likelihood* abbia le caratteristiche della *pdf* che sono qui riassunte.

4.1.1 Caratteristiche necessarie della *likelihood*

La *likelihood* è spesso funzione di molti parametri ($\vec{\theta}$ alcuni dei quali fissati, altri di cui devo trovare gli estimatori), di più variabili (\vec{x}) ed è la somma di più contributi spesso determinati in modo indipendente:

$$f(\vec{x}|\vec{\theta}) = \sum_c \xi_c f_c(\vec{x}|\vec{\theta}) \quad (4.1)$$

Per esempio quando si ha un segnale che si vuole distinguere da due sorgenti di fondo e misurarne un'asimmetria A nello spettro angolare, i contributi sono tre (uno per il segnale e due per il fondo), ci sono varie variabili discriminanti (\vec{x}), ed il parametro $\theta = A$.

Sono qui elencate le regole che bisogna seguire (e dunque le trappole più frequenti):

- le ξ_c rappresentano le frazioni di eventi che appartengono ad ogni categoria ($\sum_c \xi_c = 1$). Faremo inizialmente l'ipotesi che esse siano note salvo rilasciarla nel paragrafo 4.1.3.

- ogni f_c deve essere normalizzata ad 1, se integrata su tutte le variabili da cui dipende (non sui parametri, però). Questa normalizzazione deve essere valida per ogni valore dei parametri, il che porta tipicamente le normalizzazioni a dipendere dai parametri stessi.
- E' talvolta conveniente esprimere le f_c come funzioni di combinazioni differenti di parametri. In questi casi però bisogna assicurarsi, introducendo gli opportuni Jacobiani, di normalizzare sotto integrazione delle stesse variabili. Supponiamo per esempio che la categoria 1 abbia una likelihood $f_1(x_1, x_2, A)$ mentre la categoria 2 $f_2(x_1 + x_2, x_1 - x_2, A)$. In questo caso bisogna dividere f_2 per lo Jacobiano

$$\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = 2 \quad (4.2)$$

- le singole f_c sono spesso scritte come prodotti di *pdf* di singola variabile

$$f_c(\vec{x}) = \prod_i f_c^i(x_i) \quad (4.3)$$

Perchè questo sia possibile occorre che le variabili siano scorrelate e che le singole f_c^i siano appropriatamente normalizzate. E' da notare che spesso nei singoli contributi (c) le variabili tra loro scorrelate sono diverse, fatto che spesso porta ad una scelta diversa dell'insieme di variabili tra i vari contributi (vedi punto precedente).

4.1.2 Estimatori ottenuti con la likelihood

Consideriamo il caso di N misure indipendenti di una singola variabile ($x_j, j = 1, N$) e di un singolo parametro (θ). E' possibile costruire un estimatore di θ massimizzando, al variare di θ , la likelihood complessiva dell'insieme di misure

$$\mathcal{L}(\theta) = \prod_j f(x_j|\theta) \quad (4.4)$$

$$\hat{\theta} = \max_{\theta} \mathcal{L}(\theta) = \max_{\theta} -\log(\mathcal{L}(\theta)) = \max_{\theta} L(\theta)$$

dove si è definita la 'loglikelihood'

$$L = -\log(\mathcal{L}) = \sum_j -\log(f(x_j|\theta)) \quad (4.5)$$

Per studiare le proprietà dell'estimatore così definito consideriamo lo sviluppo in serie di f attorno al valore vero θ_T di θ :

$$\log f(x_j|\theta) = \log f(x_j|\theta_T) + \delta\theta\partial_\theta\log f + \frac{(\delta\theta)^2}{2}\partial_{\theta\theta}\log f \quad (4.6)$$

dove le derivate parziali prima $\partial_\theta f$ e seconda $\partial_{\theta\theta} f$ rispetto a θ sono sempre pensate calcolate nel valore vero θ_T .

Si ha dunque

$$L = E(x_1, \dots, x_N, \theta_T) + F(x_1, \dots, x_N, \theta_T)\delta\theta + \frac{(\delta\theta)^2}{2}G(x_1, \dots, x_N, \theta_T) \quad (4.7)$$

dove si è definito

$$\begin{aligned} E &= \sum_j (-\log f(x_j|\theta_T)) \\ F &= \sum_j \partial_\theta(-\log f(x_j|\theta_T)) \\ G &= \sum_j \partial_{\theta\theta}(-\log f(x_j|\theta_T)) \end{aligned} \quad (4.8)$$

E' anche utile definire, per F e G , le medie su tutti i possibili valori assumibili da tutte le misure:

$$\begin{aligned} \bar{F} = \frac{\langle F \rangle}{N} &= \frac{1}{N} \int (\prod_j f(x_j) dx_j) F(x_1, \dots, x_N, \theta_T) \\ &= \int dx f(x|\theta_T) \partial_\theta(-\log f(x|\theta_T)) \\ \bar{G} &= \int dx f(x|\theta_T) \partial_{\theta\theta}(-\log f(x|\theta_T)) \end{aligned} \quad (4.9)$$

e calcolando \bar{F} si trova

$$\bar{F} = \int dx f \partial_\theta(-\log f) = \partial_\theta \int f dx = 0 \quad (4.10)$$

E' infine utile sostituire le equazioni 4.9 nelle equazioni 4.8, ottenendo

$$\begin{aligned} F &= \sum_j [\partial_\theta(-\log f(x_j|\theta_T)) - \bar{F}] = \sum_j \delta F_j \\ G &= N\bar{G} + \sum_j [\partial_{\theta\theta}(-\log f(x_j|\theta_T)) - \bar{G}] = N\bar{G} + \sum_j \delta G_j \end{aligned} \quad (4.11)$$

Con queste definizioni si può facilmente ricavare il bias medio dell'estimatore ($\langle \delta\theta \rangle$) e la sua varianza. Infatti $\hat{\theta}$ si ottiene minimizzando L in funzione di θ , ossia, per l'equazione 4.5,

imponendo

$$F(x_1, \dots, x_N, \theta_T) + (\delta\theta)G(x_1, \dots, x_N, \theta_T) = 0 \quad (4.12)$$

ovverosia

$$\delta\theta = -\frac{F(x_1, \dots, x_N, \theta_T)}{G(x_1, \dots, x_N, \theta_T)} \quad (4.13)$$

Il bias si calcola calcolando la media di $\delta\theta$ su tutti i possibili valori di x_1, \dots, x_N

$$\begin{aligned} \langle \delta\theta \rangle &= -\left\langle \frac{F(x_1, \dots, x_N, \theta_T)}{G(x_1, \dots, x_N, \theta_T)} \right\rangle \\ &= -\left\langle \frac{\sum_j \delta F_j}{N\bar{G} + \sum_j \delta G_j} \right\rangle \\ &\sim -\left\langle \frac{1}{N\bar{G}} \sum_j \delta F_j \left(1 - \frac{\sum_j \delta G_j}{N\bar{G}}\right) \right\rangle \end{aligned} \quad (4.14)$$

e tende a zero per infinite misure (a meno di comportamenti anomali della f).

Dal momento che $\langle \delta F_j \rangle = 0$

$$\langle \delta\theta \rangle = \frac{\langle FG \rangle}{N\bar{G}^2} \quad (4.15)$$

$$= \frac{\int x f \partial_\theta(\log f) \partial_{\theta\theta}(\log f)}{N(\int x f \partial_{\theta\theta}(\log f))^2} \quad (4.16)$$

La varianza di $\delta\theta$, che rappresenta la risoluzione su $\hat{\theta}$ si calcola partendo da

$$\begin{aligned} E[\delta\theta^2] &= \left\langle \frac{F(x_1, \dots, x_N, \theta_T)^2}{G(x_1, \dots, x_N, \theta_T)^2} \right\rangle \\ &= \frac{\langle \sum_j \delta F_j \rangle}{(N\bar{G})^2} \\ &= \frac{\int dx f (\partial_\theta \log f)^2}{N(\bar{G})^2} \end{aligned} \quad (4.17)$$

Ma poichè

$$\begin{aligned} \bar{G} &= \int dx f \partial_\theta \frac{\partial_\theta f}{f} \\ &= -\int dx f \frac{(\partial_\theta f)^2}{f^2} + \partial_{\theta\theta} \int dx f \\ &= -\int dx f (\partial_\theta \log f)^2 \end{aligned} \quad (4.18)$$

allora si può scrivere

$$E[\delta\theta^2] = -\frac{1}{NG} \quad (4.19)$$

e nell'ipotesi di bias trascurabile l'errore su $\hat{\theta}$ è

$$\sigma_\theta = -N\partial_{\theta\theta} \int dx(-\log f) \quad (4.20)$$

ed è comunemente approssimata da

$$\sigma_\theta = -\partial_{\theta\theta} L(\theta) \quad (4.21)$$

(con L definita in equazione 4.5).

E' spesso utile trovare la varianza del parametro misurato in modo grafico. A tal fine si grafica la likelihood in funzione del parametro θ e si considera l'equazione 4.5 al primo ordine dello sviluppo in equazione 4.11:

$$L \sim E(x_1, \dots, x_N, \theta_T) + \frac{(\delta\theta)^2}{2} \bar{G} = L_{min} + \frac{(\delta\theta)^2}{2\sigma_\theta^2} \quad (4.22)$$

Si può perciò vedere che σ_θ si ottiene guardando di quanto ci si deve discostare dal valore di θ in cui si ha il minimo perchè la -loglikelihood aumenti di 0.5 rispetto al minimo L_{min} . In generale questo intervallo può anche essere asimmetrico.

Nel caso a più parametri $\vec{\theta}$ le osservazioni di questo paragrafo si possono generalizzare in

$$\begin{aligned} \widehat{\vec{\theta}} &= \max_{\vec{\theta}} \mathcal{L}(\vec{\theta}) \\ (\widehat{V_{ij}})^{-1} &= \partial_{\theta_i \theta_j} L \end{aligned} \quad (4.23)$$

La tecnica grafica continua a funzionare, ma la quota non è più 0.5, ma va cercata nelle tabelle del Particle Data Book [?]. E' invece sempre valida se si mostra il valore di L al variare di una delle variabili (θ_j) avendo minimizzato in funzione di tutte le altre variabili.

Esempio: media di una guassiana

Consideriamo il caso di eventi che provengono da una distribuzione gaussiana $f = \frac{e^{-0.5(\frac{x-\mu}{\sigma})^2}}{\sqrt{2\pi}\sigma}$ di cui si intende misurare μ . In questo caso

$$\begin{aligned} \partial_\mu(-\log f) &= \frac{x - \mu}{\sigma^2} \\ \partial_{\mu\mu}(-\log f) &= \frac{-1}{\sigma^2} \end{aligned} \quad (4.24)$$

per cui

$$\begin{aligned} \langle \delta\mu \rangle &\propto \int dx f(x - \mu) = 0 \\ \frac{1}{\sigma_\mu} &= -N \int dx f \frac{-1}{\sigma^2} = \frac{N}{\sigma^2} \end{aligned} \quad (4.25)$$

E' dunque possibile determinare il valor medio senza bias e con errore $\frac{\sigma}{\sqrt{N}}$.

4.1.3 Likelihood estesa

4.1.4 Esempi

Misura di asimmetria: $\sin 2\beta$

Un modo per studiare la violazione della simmetria discreta carica-parità (CP) è quello di produrre mesoni $Y(4S)$ che decadono in coppie di mesoni B neutri e di studiare gli eventi in cui uno dei due decade in un autostato di CP f (per esempio $f = J/\Psi K_s^0$). Sperimentalmente si è in grado di determinare se prima di decadere in f il mesone era originariamente B^0 oppure \bar{B}^0 . Questa determinazione si chiama “tagging” ed ha in generale una probabilità w di essere sbagliata. E' inoltre possibile determinare la distanza tra il punto in cui è decaduto il mesone che ha prodotto lo stato finale f ed il punto di decadimento dell'altro mesone prodotto (Δt). Questa misura non è perfetta, ma avviene con una risoluzione descritta dalla *pdf* della quantità misurata dato il valore vero $R(\Delta t - \Delta t_T | \vec{\theta})$, detta funzione di risoluzione. $\vec{\theta}$ sono i parametri da cui essa dipende. Un caso semplificato, ma non completamente distante dalla realtà è quello in cui $R(\xi|\sigma)$ è una gaussiana con valore medio nullo e varianza σ .

Nell'ipotesi in cui è possibile determinare in modo certo sia il tagging sia Δt , la *pdf* di Δt è

$$P_{s,\eta_t}(\Delta t) = N e^{-\Gamma \Delta t} (1 + \eta_t \sin 2\beta \sin \Delta M \Delta t) \quad (4.26)$$

dove il segno η_t si riferisce al fatto che il mesone decaduto è un B^0 ($\eta_t = 1$) oppure un \bar{B}^0 rispettivamente ($\eta_t = -1$), N è un fattore di normalizzazione (che non verrà esplicitato per comodità, ma che è fondamentale calcolare correttamente), Γ è la larghezza totale del B^0 , β è un parametro del Modello Standard che misura l'entità della violazione di CP e ΔM è la frequenza di mixing.

E' dunque possibile estrarre, tramite un fit di likelihood, il parametro β . In assenza di effetti del rivelatore, la likelihood sarebbe quella mostrata in equazione 4.26. Occorre però tenere conto del fatto che:

- il tagging è incorretto in una frazione w dei casi. Il che vuol dire che se il risultato del tagging è $\eta_m = \pm 1$ la pdf è una combinazione di quella con il vero $\eta_t = +1$ e -1 :

$$\begin{aligned} P_{s,\eta_m}^{tag}(\Delta t) &= (1-w)P_{s,\eta_t=\eta_m}(\Delta t) + wP_{s,\eta_t=-\eta_m}(\Delta t) \quad (4.27) \\ &= Ne^{-\Gamma\Delta t}(1 + (1-2w)\eta_m \sin 2\beta \sin \Delta M\Delta t) \end{aligned}$$

- la misura di Δt ha una risoluzione finita. Questo si esprime nella likelihood tramite una convoluzione:

$$P_{s,\eta_t}^{ris}(\Delta t) = N' \int dt P_{s,\eta_t}(t) R(\Delta t - t | \vec{\theta}) \quad (4.28)$$

E' da notare come anche qui la normalizzazione N' è cruciale.

- il campione considerato include anche eventi di fondo, la cui quantità e la cui pdf si possono ottenere da un campione di dati indipendenti (“sidebands”). Questo campione si distingue dal segnale perchè esiste una variabile (m_{es}) che ha una pdf differente per segnale e fondo. La likelihood può tenere conto di questo introducendo un nuovo termine $P_{b,\eta_m}(\Delta t | \theta_B)$ dove θ_B è un insieme di parametri che caratterizzano empiricamente il fondo e che vanno considerati come parametri rispetto ai quali minimizzare. La probabilità che un evento appartenga al segnale al variare di m_{es} è indicata come $p_s(m_{es})$.
- esistono eventi in cui lo stato finale f' è diverso, che hanno una pdf nota, ma che sono affetti dalla stessa probabilità di sbagliare tagging w e dalla stessa funzione di risoluzione:

$$P_{f',\eta_m}(\Delta t) = M \int dt R(\Delta t - t | \vec{\theta}) e^{-\Gamma t} [1 + (1-2w)\eta_m \cos \Delta M\Delta t] \quad (4.29)$$

dove M è un ulteriore fattore di normalizzazione. Questo campione di eventi può essere utilizzato per determinare direttamente i parametri w e $\vec{\theta}$ dai dati.

In definitiva dunque, si misura η_m e Δt sia in eventi $B \rightarrow f$ che $B \rightarrow f'$ e si calcola la likelihood come definita in 4.5

$$\begin{aligned} L(\beta, w, \vec{\theta}, \theta_B) &= \sum_{B \rightarrow f} -\log(f(\Delta t, \eta_m | \beta, w, \vec{\theta}, \theta_B)) \quad (4.30) \\ &+ \sum_{B \rightarrow f'} -\log(f'(\Delta t, \eta_m | w, \vec{\theta})) \end{aligned}$$

dove

$$f(\Delta t, \eta_m | \beta, w, \vec{\theta}, \theta_B) = N [p_s(m_{es}) \int dt R(\Delta t - t | \vec{\theta}) e^{-\Gamma t} \dots] \quad (4.31)$$

$$\begin{aligned}
& [1 + (1 - 2w)\eta_t \sin 2\beta \sin \Delta Mt] \\
& + (1 - p_s(m_{es}))P_{b,\eta_m}(\Delta t|\theta_B)], \\
f'(\Delta t, \eta_m|w, \vec{\theta}) &= P_{f',\eta_m}(\Delta t)
\end{aligned}$$

La likelihood viene dunque minimizzata al variare di tutti i parametri da cui dipende. Si calcola la matrice di covarianza in base all'equazione 4.23 e si ottiene così il risultato per $\sin 2\beta$.

Da questo esempio si impara che:

- bisogna stare attenti alle normalizzazioni
- è possibile tenere conto di probabilità di errore, funzioni di risoluzione e fondo direttamente nello scrivere la likelihood (“folding”)
- quando possibile è sempre consigliabile ricavare i parametri che caratterizzano la risposta del rivelatore dai dati. Questo solitamente si effettua calcolando la likelihood sulla somma del campione dei dati e dei campioni “di controllo”, capaci di portare informazioni sui parametri incogniti.

Conteggio numero di eventi:misura di Branching Fraction in presenza di fondo

Un'applicazione leggermente diversa si ha quando si vuole misurare quantità legate a numeri di eventi prodotti, per esempio i Branching Fraction $B \rightarrow \pi^+\pi^-$ e $B \rightarrow K^+\pi^-$. Questi eventi sono tipicamente confusi tra copiosi quantitativi di fondo e non esiste una singola variabile che sia capace di separare segnale e fondo in modo efficiente e puro, ma esistono un numero di variabili \vec{x} distribuite in modo diverso tra segnale e fondo. Se dunque si indicano con $P_{\pi\pi}(\vec{x})$, $P_{K\pi}(\vec{x})$ e $P_B(\vec{x})$ le *pdf* delle variabili discriminanti nei due segnali di interesse e nel fondo, è possibile scrivere l'extended likelihood come

$$e^{-L}(N_{\pi\pi}, N_{K\pi}, N_B) = e^{-N'}(N')^N \prod_{i=1}^N P_i(N_{\pi\pi}, N_{K\pi}, N_B) \quad (4.32)$$

dove $N' = N_{\pi\pi} + N_{K\pi} + N_B$, N è il numero totale di eventi osservati, la produttoria si estende a tutti gli eventi e la probabilità di singolo evento è

$$P_i(N_{\pi\pi}, N_{K\pi}, N_B) = \frac{N_{\pi\pi}P_{\pi\pi}(\vec{x}_i^\dagger) + N_{K\pi}P_{K\pi}(\vec{x}_i^\dagger) + N_B P_B(\vec{x}_i^\dagger)}{N'} \quad (4.33)$$

Si trova dunque il minimo di L e dunque il numero di eventi di segnale (e di conseguenza si può calcolare il Branching Fraction).

La parte delicata di questa misura è la scrittura delle *pdf*. Esse vengono infatti spesso costruite come prodotti di funzioni delle singole variabili (per esempio $P_{\pi\pi}(\vec{x}) = \prod_k f_{\pi\pi}^k(x_k)$), ma questo è valido solo nell'ipotesi che le variabili siano scorrelate. La forma delle *pdf* viene estratta da campioni di controllo nei dati quando possibile, dalla simulazione altrimenti. L'ipotesi che le correlazioni tra le variabili sia trascurabile viene testata con la simulazione, salvo poi stimare opportunamente le incertezze derivanti da possibili imprecisioni nella simulazione.

4.2 χ^2 e minimi quadrati

Nel caso in cui le misure siano distribuite in modo gaussiano, la likelihood si riduce a

$$\begin{aligned} L(\vec{\theta}) &= \frac{1}{2}(\vec{x} - \vec{g}(\vec{\theta}))\widehat{V}^{-1}(\vec{x} - \vec{g}(\vec{\theta})) \\ &= \frac{1}{2}\chi^2(\vec{\theta}) \end{aligned} \quad (4.34)$$

dove g_i è il valore di aspettazione della variabile x_i ed è funzione dei parametri che si intende misurare, $\vec{\theta}$.

Questo caso speciale della likelihood è noto come fit del χ^2 ovvero come metodo dei minimi quadrati. Le sue caratteristiche non sono però altro che applicazioni delle formule già viste nel caso della likelihood:

- l'estimatore di $\vec{\theta}$ si ottiene minimizzando $\chi^2(\vec{\theta})$.
- la matrice di covarianza dei parametri è legata all'Hessiano del χ^2 :

$$\widehat{V}_{ij}^{-1} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j} \quad (4.35)$$

- graficamente si può ricavare l'intervallo di confidenza di un parametro guardando alla dipendenza del χ^2 dal parametro stesso e considerando l'intervallo in cui il χ^2 dista di meno di 1 dal valore di minimo come intervallo di confidenza al 68% C.L.

4.2.1 Media pesata

Un'applicazione molto usata del metodo dei minimi quadrati è la determinazione della media pesata di N misure x_i della stessa grandezza fisica, con distribuzione gaussiana e matrice di covarianza nota

V_{ij} . In questo caso l'ipotesi che si fa è che tutte le misure abbiano lo stesso valore di aspettazione x_0 per cui

$$\chi^2(x_0) = \sum_{ij} (x_i - x_0)(V^{-1})_{ij}(x_j - x_0) \quad (4.36)$$

ed il miglior estimatore di x_0 si ottiene imponendo a 0 la derivata del χ^2

$$\widehat{x}_0 = \frac{\sum_{ij} x_i (V^{-1})_{ij}}{\sum_{ij} (V^{-1})_{ij}} \quad (4.37)$$

L'errore ottenuto dalla derivata seconda è

$$\widehat{\sigma}_{x_0}^2 = \frac{1}{\sum_{ij} (V^{-1})_{ij}} \quad (4.38)$$

Nel caso in cui le misure siano da considerarsi scorrelate

$$\begin{aligned} \widehat{x}_0 &= \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}} \\ \widehat{\sigma}_{x_0}^2 &= \frac{1}{\sum_i \frac{1}{\sigma_i^2}} \end{aligned} \quad (4.39)$$

nota come media pesata con pesi $w_i = \frac{1}{\sigma_i^2}$

Errori correlati e scorrelati

Le equazioni 4.37 e 4.38 sono necessarie e sufficienti per combinare misure sperimentali con errori sistematici e statistici, a patto che si sappia che gli errori sono gaussiani e si sappia quale componente degli errori è correlata o scorrelata.

Mentre l'errore statistico delle misure è tipicamente scorrelato (σ_i^s), gli errori sistematici delle singole misure si dividono in errore scorrelato (σ_i^u) ed errore correlato (σ_i^c). Per errore correlato tra più misure si intende di quanto si muovono tutte insieme le misure se si è fatto un dato errore. Per esempio se si sta misurando il numero di K_s^0 prodotti e si ricostruisce il K_{0s} tramite il suo decadimento in due pioni carichi, l'incertezza sul conteggio del segnale $K_s^0 \rightarrow \pi^+\pi^-$ è scorrelata, mentre l'incertezza sul Branching Fraction $BR(K_s^0 \rightarrow \pi^+\pi^-)$ è correlata. Occorre notare che la separazione in errore correlato e scorrelato dipende dall'insieme delle misure che si considera e va riconsiderata se si cambia questo insieme.

Si può scrivere la matrice di covarianza come somma di tre termini:

$$V_{ij} = V_{ij}^s + V_{ij}^u + V_{ij}^c \quad (4.40)$$

dove il termine statistico e quello scorrelato sono, per definizione di matrice di covarianza

$$V_{ij}^s = (\sigma_i^s)^2 \delta_{ij} \quad (4.41)$$

$$V_{ij}^u = (\sigma_i^u)^2 \delta_{ij} \quad (4.42)$$

mentre il termine correlato è

$$V_{ij}^c = \sigma_i^c \sigma_j^c \quad (4.43)$$

visto che per la proprietà del coefficiente di correlazione descritte nel cap 2.4 in questo caso tutti i coefficienti di correlazione sono unitario.

Data la covarianza nell'equazione 4.40, è possibile calcolare la media pesata. Va peraltro notato che V_{ij}^c non è invertibile, ma questo è corretto perchè se gli unici errori fossero tutti correlati, le misure sarebbero legate tra loro da relazioni funzionali e dunque data una misura tutte le altre sarebbero forzate ad avere un valore assegnato ed il χ^2 tra loro è per definizione infinito.

4.2.2 Esempio: Fit dei dati ai parametri del Modello Standard

4.3 Fit ad istogrammi

Fin qui si sono trattati fit in cui ogni evento viene trattato individualmente. Essi sono più accurati e sono i più utilizzati per ottenere risultati attendibili. Visto però che spesso si ricorre a minimizzazioni numeriche e queste sono talvolta troppo pesanti dal punto di vista computazionale, è spesso utile fittare direttamente istogrammi, almeno per avere stime approssimate.

Come descritto in 2.2.2, un istogramma è

4.4 Esercizi

Esercizio 12- Fit di asimmetria avanti-indietro

La direzione angolare di un quark prodotto dal decadimento del bosone Z è

$$f(\theta) = (1 + \cos^2\theta + A\cos\theta)$$

Si sono effettuate N misure (θ_i) con risoluzione σ_θ . L'efficienza di rivelazione è $\epsilon(\theta)$. Scrivere la Likelihood e trovare una espressione approssimata per A nel caso in cui non si tenga in conto la risoluzione.

Esercizio 11- Risoluzione in impulso

La risoluzione sulla misura dell'impulso trasverso di una traccia

$$\text{è: } r = \frac{\sigma_{p_T}}{p_T} = \sqrt{a + bp_T^2}$$

Abbiamo diviso N tracce in n bin di p_T e misurato perciò n coppie (r_i, p_T^i) con $V(r_i) = \sigma_i^2$. Determinare a e b .

Capitolo 5

Test di ipotesi

Molto spesso è necessario verificare se un campione di dati soddisfa o meno delle assunzioni, dette "ipotesi statistiche". Per esempio, se si vuole decidere se una data moneta è truccata, si formula l'ipotesi che essa non lo sia, $p = 0.5$, in cui p è la probabilità che si presenti testa. In modo simile se si intende decidere se una procedura è migliore di un'altra, formuleremo l'ipotesi che non c'è differenza fra le due procedure (ovvero la differenze osservate siano mere fluttuazioni dovute al campionamento). Questo tipo di ipotesi è spesso chiamata *Ipotesi nulla* e denotata da H_0 .

Qualunque ipotesi che differisca da un'ipotesi data è detta *Ipotesi alternativa*. Ad esempio, se un'ipotesi è $p = 0.5$, le ipotesi $p = 0.7$ o $p > 0.5$ sono alternative. Un'ipotesi alternativa di un'ipotesi nulla è denotata da H_1 .

Se si trova che la probabilità che i risultati osservati su un campione differiscano da quelli che, per puro effetto del caso, ci si sarebbe aspettati sulla base dell'ipotesi formulata, è maggiore di una soglia predeterminata, le differenze osservate sono significative e l'ipotesi viene rigettata.

Le procedure che permettono di decidere se accettare o rigettare una data ipotesi sono dette *Test d'ipotesi*.

Se un'ipotesi viene rigettata quando avrebbe dovuto essere accettata, si parla di *Errore di I tipo*. Se d'altro canto, un'ipotesi viene accettata quando avrebbe dovuto essere rigettata, si parla di *errore di II tipo*. La scelta del test d'ipotesi da utilizzare viene effettuata minimizzando gli errori di decisione. Ciò non è un fatto semplice poiché, per una data grandezza del campione, i tentativi di diminuire gli errori di un tipo sono in genere accompagnati da un aumento di errori dell'altro tipo. Il solo modo di ridurre entrambi gli errori è di aumentare le dimensioni del campione, operazione che a volte può essere possibile, a volte no.

Nell'operare un test d'ipotesi, la massima probabilità con la quale siamo disposti a rischiare un errore del I tipo è detta *Livello di significatività* del test. Questa probabilità è denotata spesso con α . In pratica si utilizzano livelli di significatività del 5% o 10%, cioè siamo fiduciosi al 95% o al 90% di aver preso la decisione giusta.

5.1 Distribuzione del chiquadro: χ^2

La funzione χ^2 introdotta nel paragrafo 4.2 può essere pensata come una variabile casuale. Le sue proprietà sono note ed è dunque possibile estrarre importanti informazioni da esse.

Consideriamo un set di variabili casuali mutuamente indipendenti $y_1 \dots y_n$ tutte gaussiane

$$f(y_i) = g(\mu = 0, \sigma = 1)$$

Definiamo χ^2 la variabile

$$\chi_n^2 = \sum_{i=1}^n y_i^2 \quad (5.1)$$

ed n gradi di libertà della variabile. La pdf della variabile χ^2 è nota se conosciamo il numero gradi di libertà:

$$f(\chi_n^2, n) = C(\chi_n^2, n) \quad (5.2)$$

Essa è tale che

$$\begin{aligned} E[\chi^2 | n] &= n \\ V[\chi^2 | n] &= 2n \end{aligned} \quad (5.3)$$

Per illustrare il test di ipotesi conviene considerare il caso di variabili correlate. Date N misure x_i , $i = 1 \dots N$ ed un insieme di parametri \vec{g} , se i valori di aspettazione delle misure sono

$$E[x_i] = \phi_i(\vec{g}) \quad (5.4)$$

e la matrice di correlazione

$$V_{ij}[\vec{x}] = V_{ij} \quad (5.5)$$

è possibile testare l'ipotesi che le funzioni ϕ_i siano la corretta espressione dei valori di aspettazione, oppure che \vec{g} sia il valore corretto dei parametri. Si calcola infatti

$$\chi^2 = (\vec{x} - \vec{\phi}(\vec{g}))V^{-1}(\vec{x} - \vec{\phi}(\vec{g})) = -2 \ln G(\vec{x} | \vec{\mu} = \vec{\phi}; \hat{V} = \hat{v}) + c \quad (5.6)$$

ed il numero di gradi di libertà $\nu = N$. E' dunque possibile calcolare la probabilità di ottenere un χ^2 maggiore di quello misurato $C(\chi^2, n)$ e questo valore è maggiore di una soglia predefinita α allora il test è superato con il livello di confidenza α .

Nel caso in cui M dei parametri $\vec{\gamma}$ sono stati stimati sullo stesso campione di dati, allora il numero di gradi di libertà va ricalcolato come $\nu = N - M$.

5.1.1 Esempi

Combinazione di misure

Nella sessione 4.2.1 abbiamo visto come sia possibile combinare misure della stessa quantità a patto che siano distribuite in modo gaussiano e siano noti gli errori correlati e scorrelati.

Sotto la stessa ipotesi è anche possibile stimare il livello di confidenza che queste misure stiano effettivamente misurando le stesse cose o che gli errori e le correlazioni siano corrette.

A tal fine basta calcolare il χ^2 in equazione 4.36 avendo sostituito a x_0 il suo valore nel minimo in equazione 4.37. $C(\chi^2, N - 1)$ rappresenta il livello di confidenza richiesto.

Ricostruzione vertici di decadimento

Supponiamo di conoscere la traiettoria di N tracce con errore sui suoi parametri, e di voler stimare la probabilità che provengano dallo stesso vertice e stimare la posizione del vertice comune.

Supponiamo di poter risalire, dai parametri della traiettoria ad N punti di minimo avvicinamento ad un possibile candidato vertice \vec{x}_v e supponiamo di conoscere la posizione (\vec{x}_i), l'impulso (\vec{p}_i) e la matrice di covarianza ($\widehat{V}_{xx}^i, \widehat{V}_{pp}^i, \widehat{V}_{xp}^i$) in questi punti.

Il χ^2 si scrive allora

$$\begin{aligned} \chi^2(\vec{x}_v, p_v^1, \dots, p_v^N) &= \sum_i (\vec{x}_v - \vec{x}_i) \widehat{V}_{xx}^i (\vec{x}_v - \vec{x}_i) \\ &+ (\vec{p}_v^i - \vec{p}_i) \widehat{V}_{pp}^i (\vec{p}_v^i - \vec{p}_i) \\ &+ (\vec{x}_v - \vec{x}_i) \widehat{V}_{xp}^i (\vec{p}_v^i - \vec{p}_i) \end{aligned} \quad (5.7)$$

dove p_v^i è l'impulso in \vec{x}_v dalla particella i -esima (si sta facendo un'ipotesi lineare sulla correlazione tra impulso e posizione).

I parametri del χ^2 vengono dunque lasciati liberi di variare e si cerca per quale valore il χ^2 assume il minimo valore. Tipicamente è

necessario iterare il processo ed usare il valore misurato come punto di partenza per la nuova iterazione.

Per calcolare la probabilità che tutte le tracce provenngano dallo stesso vertice si utilizza il minimo χ_{vert}^2 , avendo calcolato correttamente il numero di gradi di libertà . Prima di minimizzare, i gradi di libertà sono $6n$, con n tracce nel vertice, cioè tante quante sono le gaussiane in $\chi^2(\vec{x}_v, \vec{p}_v)$. Dopo la minimizzazione, i gradi di libertà sono $6n$ meno il numero di parametri che sono stati fissati: $\nu = 6n - 3(n - 1) = 3(n + 1)$.

5.2 Test di student

5.3 Test di Kolmogorof

5.4 Esercizi

Esercizio 14- test di student In riferimento all'esercizio 2, nell'ipotesi che tutti gli errori siano gaussiani e scorrelati, determinare la probabilità che le misure di SLD e LEP siano consistenti.

Capitolo 6

Metodi MonteCarlo

INTRODUZIONE SULL'USO DEL MC

6.1 Metodi Montecarlo

6.1.1 Generazione di numeri casuali uniformi

6.1.2 Metodo di Wilson

6.1.3 Metodo diretto

6.1.4 Hit & Miss

6.1.5 Campionamento di importanza

6.1.6 Metodo dei pesi

6.2 Esempio in fisica delle alte energie

Capitolo 7

Estrazione di limiti

7.1 Conteggio di eventi

7.1.1 limiti in assenza di fondo

7.1.2 limiti in presenza di fondo

7.2 limiti con likelihood

7.3 Feldman Cousin

7.4 Esercizi

Esercizio 15-Poisson

Dopo aver raccolto una luminosità integrata L trovare la più alta sezione d'urto di un processo compatibile coll'osservazione di 0 eventi al 95% C.L.. Sia ϵ l'efficienza di rivelazione di questa specie di eventi.

Esercizio 16- I 4 Jet di Aleph

In un istogramma con 20 bin che sono previsti avere lo stesso numero di entrate per bin ($\mu = 0.5$), due dei bin hanno in totale 10 eventi. Calcolare la probabilità che questo sia solo una fluttuazione statistica e non un segnale di fisica.

Esercizio 17

Si è raccolta una luminosità integrata $L = \int L dt = 1 \text{ pb}^{-1}$. Il processo che vogliamo misurare (segnale) ha una sezione d'urto effettiva $\sigma_S \epsilon_s = 4 \text{ pb}$, mentre i processi di fondo hanno $\sigma_B \epsilon_B = 1 \text{ pb}$. Quale è il massimo numero di eventi che potete aver osservato per poter escludere la presenza di un segnale al 95% C.L.

Esercizio 18

Con una luminosità integrata di un picobarn inverso e una sezione d'urto per il fondo $\sigma_B = 1$ pb qual'è il più piccolo segnale per il quale si ha ancora sensibilità.

Capitolo 8

Soluzione degli esercizi

Esercizio 1

La media LEP è 0.231866 con $\sigma = 0.00045$. L'errore sulla media lep e' allora

$$\mu_{LEP} = 0.23187 \pm 0.00020$$

La differenza tra LEP ed SLD è

$$\mu_{LEP} - \mu_{SLD} = 0.00132 \pm 0.00043$$

Esso non è consistente con 0 entro 1σ . Poichè però non abbiamo fatto ipotesi sulla funzione di distribuzione delle variabili non è possibile fare alcuna affermazione quantitativa sulla eventualità che non entrambi gli esperimenti possono essere corretti.

Esercizio 2

L'istogramma dice quanti eventi hanno dato una data risposta i (N_i). Essendo queste misure indipendenti di una sola variabile, e' possibile stimare la media ($\hat{\mu}$) e la varianza ($\widehat{s^2}$) con una versione modificata delle equazioni 1.33 e 1.17:

$$\hat{\mu} = \frac{\sum_i i N_i}{N} = 5.53$$

$$\widehat{s^2} = \frac{\sum_i i^2 N_i}{N-1} - \frac{N \hat{\mu}^2}{N-1} = 4.02$$

$$\hat{s} = \sqrt{4.02} = 2.00$$

dove $N = \sum_i N_i$. Le incertezze su queste variabili sono

$$\sigma_{\hat{\mu}} = \sqrt{\frac{\widehat{s^2}}{N}} = 0.52$$

$$\sigma_{\hat{s}} = \sqrt{\frac{\widehat{s^2}}{2(N-1)}} = 0.38$$

Esercizio 3A) $p = \frac{n}{N}$

B) bisogna tenere in conto la probabilità combinata di vincere la seconda volta se si è persa la prima. In questo caso semplice conviene enumerare tutte le possibilità.

<i>I</i>	<i>II</i>	$P_I P_{II}$	P_T
<i>SI</i>	<i>SI</i>	$\frac{n}{N} \frac{n-1}{N-1}$	$\frac{n(n-1)}{N(N-1)}$
<i>NO</i>	<i>NO</i>	$(1 - \frac{n}{N})(1 - \frac{n}{N-1})$	$\frac{(N-n)(N-1-n)}{N(N-1)}$
<i>SI</i>	<i>NO</i>	$\frac{n}{N}(1 - \frac{n-1}{N-1})$	$\frac{n(N-n)}{N(N-1)}$
<i>NO</i>	<i>SI</i>	$(1 - \frac{n}{N})\frac{n}{N-1}$	$\frac{(N-n)n}{N(N-1)}$

Quindi la probabilità di avere almeno una vittoria sarà $P_V = 1 - \textit{NONO}$

$$\begin{aligned} P_V &= 1 - \frac{(N-n)(N-1-n)}{N(N-1)} = \frac{nN + n(N-1-n)}{N(N-1)} \\ &= \frac{n(2N-1-n)}{N(N-1)} \end{aligned}$$

e la probabilità di avere solo una vittoria

$$P_1 = 2 \frac{n(N-n)}{N(N-1)}$$

e quella di due vittorie

$$P_2 = \frac{n(n-1)}{N(N-1)}$$

Casi notevoli sono :

$$N = n \quad P_V = 1, \quad P_1 = 0, \quad P_2 = 1$$

$$n = 1 \quad P_V = \frac{2}{N}, \quad P_1 = \frac{2}{N}, \quad P_2 = 0$$

Il risultato intuitivo sarebbe $P_V = \frac{2n}{N}$. Esso equivale al limite $1 < n \ll N$ dove

$$P_V = \frac{n(2N-1-n)}{N(N-1)} = \frac{n}{N} \frac{2 - \frac{1}{N} - \frac{n}{N}}{1 - \frac{1}{N}} \simeq \frac{2n}{N}$$

Nell'approccio di Bayes (vedi capitolo 3)

$$P(SI) = \frac{n}{N}, P(NO) = \frac{N-n}{N}$$

sono le probabilità a priori.

$$P(SI|NO) = \frac{n}{N-1}, P(SI|SI) = \frac{n-1}{N-1}, P(NO|i) = 1 - P(SI|i)$$

con $i = SI, NO$.

$$P(SI, SI) = P(SI|SI)P(SI)$$

$$P(NO, SI) = P(SI|NO)P(NO)$$

Esercizio 4

A) La probabilità di avere almeno un leptone è

$$P(\neq 0) = 1 - P(0) = 1 - (1 - B_{SL})^4 \simeq 60\%$$

mentre la probabilità di trovarne esattamente uno è

$$P(1) = 4B_{SL}(1 - B_{SL})^3 \simeq 41\%$$

B) la probabilità di avere meno di due leptoni è

$$P(> 1) = 1 - (P(0) + P(1)) = P(\neq 0) - P(1) \simeq 39\%$$

C)

Calcoliamo la probabilità di avere esattamente (nD^*, ml) . Per ogni B tre cose posso succedere:

- c'è solo una D^* nel decadimento ($a = P(D^*, \bar{l}) = B_D^* - B_{D^*l} = 20\%$)
- c'è solo un leptone nel decadimento ($b = P(\bar{D}^*, l) = B_{SL} - B_{D^*l} = 15\%$)
- ci sta sia un leptone che un D^* nel decadimento ($c = P(D^*, l) = B_{D^*l} = 5\%$)

Fissato il numero di quark nell'evento in cui si verifica il terzo caso (k) il numero di volte in cui gli altri due casi si devono verificare per avere esattamente nD^* ed ml :

$$\begin{aligned} P_{n,m} &= \sum_{k=k_{min}}^{k_{max}} \binom{4}{k} \binom{4-k}{n-k} \binom{4-m}{m-k} a^{n-k} b^{m-k} c^k (1-a-b-c)^{4-n-m+k} \\ &= a^n b^m (1-a-b-c)^{4-n-m} \sum_{k=k_{min}}^{k_{max}} \binom{4}{k} \binom{4-k}{n-k} \binom{4-m}{m-k} \end{aligned} \quad (8.1)$$

dove $k_{min} = \max(0, m + n - 4)$ e $k_{max} = \min(n, m)$ e si è fatto uso del fatto che in questo caso

$$c(1 - a - b - c) = ab$$

Per cercare il massimo conviene scrivere un piccolo programmino fortran e si trova che i due casi favoriti sono quelli in cui c'è un solo D^* e nessun oppure un leptone. In questo caso la probabilità è del 17.3%.

Esercizio 5

La distribuzione congiunta delle due variabili è:

$$g(k_1, k_2) = e^{-\mu_1} \frac{\mu_1^{k_1}}{k_1!} e^{-\mu_2} \frac{\mu_2^{k_2}}{k_2!}$$

La distribuzione della somma delle variabili si ottiene sommando su tutte le combinazioni che hanno lo stesso valore di k

$$g(k) = \sum_{k_1, k_2} \delta(k_1 + k_2 - k) g(k_1, k_2) = \sum_{k_1, k_2} \delta(k_1 + k_2 - k) e^{-\mu_1} \frac{\mu_1^{k_1}}{k_1!} e^{-\mu_2} \frac{\mu_2^{k_2}}{k_2!}$$

$$= e^{-(\mu_1 + \mu_2)} \sum_{k_1} \frac{\mu_1^{k_1}}{k_1!} \frac{\mu_2^{k-k_1}}{(k-k_1)!} = e^{-(\mu_1 + \mu_2)} \sum_{k_1} \frac{\mu_1^{k_1}}{k_1!} \frac{(\mu_2)^{k-k_1}}{(k-k_1)!}$$

introducendo $p = \mu_1/\mu$ e $q = 1 - p$

$$g(k) = e^{-(\mu_1 + \mu_2)} \sum_{k_1} \mu^k \frac{p^{k_1} q^{k-k_1}}{k_1! (k-k_1)!} = e^{-\mu} \frac{\mu^k}{k!} \sum_{k_1} \frac{k!}{k_1! (k-k_1)!} p^{k_1} q^{k-k_1} = e^{-\mu} \frac{\mu^k}{k!}$$

k è distribuita in modo poissoniano con valor medio pari alla somma dei valori medi.

Esercizio 6

A)

$$P_\mu(N) = \frac{\mu^N e^{-\mu}}{N!} = 3.7 e^{-3.7} = 0.091$$

B)

L'errore su ϵ è :

$$\sigma(\epsilon) = \sqrt{\frac{0.37 \times 0.63}{1000}} = \sigma_\epsilon = 0.015$$

$$\sigma(\mu) = k \sigma_\epsilon$$

La distribuzione di probabilità di μ è:

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} e^{-\frac{1}{2} \left(\frac{\mu - k\epsilon}{\sigma_\epsilon} \right)^2}$$

per cui

$$P(N) = \int f(\mu) P_\mu(N) d\mu = \frac{1}{\sqrt{2\pi}\sigma_\epsilon} \int e^{-\frac{1}{2} \left(\frac{\mu - k\epsilon}{\sigma_\epsilon} \right)^2} \frac{\mu^N}{N!} e^{-\mu} d\mu$$

Esercizio 7

$$N_c = \int L\sigma_{bb}\epsilon BR \Rightarrow BR = \frac{100}{0.371010^6} \simeq 2.510^{-5}$$

$$\epsilon = \frac{N_{sel}(MC)}{N_{Tot}} = 0.37$$

$$\sigma(\epsilon) = \sqrt{\frac{0.37 \times 0.63}{1000}} = \sigma_\epsilon$$

N_c è distribuito in modo poissoniano attorno a μ quindi l'errore su N_c è $\sqrt{N_c}$.

$$\Rightarrow BR = \frac{N_c}{\int L\sigma_{bb}\epsilon} \pm \frac{\delta N_c}{\int L\sigma_{bb}\epsilon} \pm \frac{N_c}{\int L\sigma_{bb}\epsilon^2} \sqrt{\frac{\epsilon(1-\epsilon)}{N_{MC}^{Tot}}}$$

Esercizio 8

A)

$$P(x_1 > 0) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$P(x_2 > 0) = 1 - P(x_1 > 0)$$

quindi

$$P(x_1|x > 0) = \frac{\frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}t^2} dt}{1 - \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}t^2} dt + \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}t^2} dt} = \text{erf}\left(\frac{x-\mu}{\sigma}\right)$$

B)

La probabilità che x_F venga da un quark positivo(+) e x_B da un quark negativo (-) è

$$P(x_F, +; x_B, -) = P(x_+ - x_- > 0)$$

$$\begin{aligned} P(x_+ - x_- = \Delta) &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^\infty \int_{-\infty}^\infty \delta(x_+ - x_- - \Delta) e^{-\frac{1}{2}\left(\frac{x_+-\mu}{\sigma}\right)^2} e^{-\frac{1}{2}\left(\frac{x_--\mu}{\sigma}\right)^2} dx_+ dx_- \\ &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^\infty e^{-\frac{1}{2}\left[\left(\frac{x_+-\mu}{\sigma}\right)^2 + \left(\frac{x_+-\Delta-\mu}{\sigma}\right)^2\right]} dx_+ \end{aligned}$$

$$= \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[x_+^2 + \mu^2 + \frac{\Delta^2}{2} - \delta(x_+ + \mu)]} dx$$

poniamo $(x_+ - \frac{\Delta}{2})^2 = x_+^2 + \frac{\Delta^2}{4} - \Delta x_+ = \xi^2$

$$P(x_+ - x_- = \Delta) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[\xi^2 - \frac{\Delta^2}{4} + \frac{\Delta^2}{2} + \mu^2 - \Delta\mu]} d\xi$$

$$= e^{-\frac{1}{2\sigma^2}[\frac{\Delta^2}{4} + \mu^2 - \Delta\mu]} \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\sqrt{2}\xi)^2} d\xi$$

poiche' $\frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(\sqrt{2}\xi)^2} d\xi = \frac{1}{2\sqrt{\pi}\sigma}$

$$P(x_+ - x_- = \Delta) = \frac{1}{2\sqrt{\pi}\sigma} e^{-\frac{1}{2\sigma^2}[\frac{\Delta^2}{4} + \mu^2 - \Delta\mu]}$$

ora sostituendo $\xi = \frac{\Delta}{\sqrt{2}} - \sqrt{2}\mu$ e quindi $d\Delta = \sqrt{2}d\xi$

$$P(x_+ - x_- > 0) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\frac{\mu}{\sqrt{2}}}^{\infty} e^{-\frac{1}{2\sigma^2}\xi^2} d\xi = \text{erf}\left(-\frac{\mu}{\sqrt{2}}\sigma\right)$$

C)

Dal momento che non so distinguere tra quark positivi o negativi a priori devo usare quantità simmetriche in x_F e X_B . In particolare si può usare:

$$E(x_F x_B) = E(x_+ x_-) = \mu^2 + \rho_{+-}\sigma^2$$

$$V(x_F + x_B) = E[(x_+ - \mu + x_- + \mu)^2] = E(x_+^2 + x_-^2 - 2E(x_+ x_-))$$

tenendo conto che $E(x_+^2) = E(x_-^2) = \sigma^2 - \mu^2$ si ottengono delle equazioni risolvibili:

$$E(x_F x_B) = \mu^2 + \rho_{+-}\sigma^2$$

$$V(x_F + x_B) = 2[(1 + \rho_{+-})\sigma^2 + 2\mu^2]$$

Esercizio 9

A)

$$P(k|x_i) = \frac{P(x_i|k)P_k}{\sum_j P(x_i|j)P_j}$$

dove P_j è la probabilità a priori.

e $P(k|x_i)$ è la probabilità di avere la specie k secondo il rivelatore i -esimo.

$$P(x_i \dots x_N | k) = \prod_i P(x_i | k)$$

è la probabilità combinata di tutti i rivelatori sull'ipotesi K

$$P(k|x_1 \dots x_N) = \frac{P(x_N|k)P(k|x_1 \dots x_{N-1})}{\sum_j P_j|x_1 \dots x_{N-1}} = \frac{\prod_i P(x_i|k)P(k)}{\sum_j \prod_i P(x_i|j)P_j}$$

Allo stesso risultato si poteva arrivare dalla probabilità totale.

B)

La probabilità che se la particella è π ha rilasciato nel rivelatore 1 un segnale ξ_1 è:

$$P(\xi_1|\pi) = \int_{\xi_1}^{\infty} g(x, 0, 1) dx$$

$$P(\xi_1|k) = \int_{-\infty}^{\xi_1} g(x, 2, 1) dx$$

con $g(x, \mu, \sigma)$ = gaussiana di media μ ed errore σ . Applicando il teorema di Bayes (probabilità a priori uguale a 1):

$$P(\pi|\xi_2) = \frac{P(\xi_2|\pi)}{P(\xi_2|\pi) + P(\xi_1|k)}$$

analogamente

$$P(\pi|\xi_2) = \frac{\int_{\xi_2}^{\infty} g(x, 0, 1) dx}{\int_{\xi_2}^{\infty} g(x, 0, 1) dx + \int_{-\infty}^{\xi_2} g(x, 10, 1) dx}$$

Per calcolare $P(\pi|\xi_1, \xi_2)$ devo applicare la composizione dei livelli di confidenza:

$$P(\pi|\xi_1, \xi_2) = P(\pi|\xi_1)P(\pi|\xi_2)(1 - \log P(\pi|\xi_1)P(\pi|\xi_2))$$

Considerare il caso particolare in cui $\xi_1 = \xi_2 = 1$.

In questo caso:

$$P(\pi|\xi_1) = \frac{1}{2}; (P(\xi_1|\pi) = P(\xi_1|k))$$

$$P(\pi|\xi_2) = \frac{0.16}{0.16 + \int_1^{\infty} g(1, 10, 1) dx} = \frac{0.16}{0.16 + \epsilon} = 1 - \frac{\epsilon}{0.16} \simeq 1$$

Il secondo rivelatore sa discriminare con grande precisione mentre il primo non ha alcun potere discriminante. Se si usasse come probabilità combinata il prodotto:

$$P_T = P(\pi|\xi_1)P(\pi|\xi_2) = 0.5(1 - \frac{\epsilon}{0.16}) \simeq 0.5$$

e non si avrebbe discriminazione. Con la formula corretta : $P = P_T(1 - \log P_T) = 0.5(1 - \log 0.5) = 0.85$

In generale se uno dei rivelatori è certo di una ipotesi mentre l'altro ha una probabilità p $P = p(1 - \log p)$ e p deve essere MOLTO piccola per ridurre significativamente P .

Esercizio 11

$$r_i^2 = a + bP_T^2, V(r_i^2) = 4r_i^2\sigma_i^2$$

$$\chi^2 = \sum_i \frac{(r_i^2 - a - bp_T^2)^2}{4r_i^2\sigma_i^2}$$

derivando rispetto ad a e b otteniamo due equazioni:

$$\sum_i \frac{(r_i^2 - a - bp_T^2)}{r_i^2\sigma_i^2} = 0$$

$$\sum_i \frac{p_i^2(r_i^2 - a - bp_T^2)}{r_i^2\sigma_i^2} = 0$$

in forma simbolica

$$M_{aa}a + M_{ab}b = N_a$$

$$M_{ba}a + M_{bb}b = N_b$$

con

$$M_{aa} = \sum_i \frac{1}{r_i^2\sigma_i^2}, M_{ab} = \sum_i \frac{p_i^2}{r_i^2\sigma_i^2}, N_a = \sum_i \frac{1}{\sigma_i^2}$$

$$M_{ba} = \sum_i \frac{p_i^2}{r_i^2\sigma_i^2}, M_{bb} = \sum_i \frac{p_i^4}{r_i^2\sigma_i^2}, N_b = \sum_i \frac{P_i^2}{\sigma_i^2}$$

discutere il caso in cui anche p_T abbia un errore.

Esercizio 12

$$L = \prod_i \int (1 + \cos^2\theta + A\cos\theta) \frac{1}{\sqrt{2\pi}\sigma_\theta} e^{-\frac{1}{2}\left(\frac{\theta-\theta_i}{\sigma_\theta}\right)^2} \epsilon(\theta) d\theta$$

L'espressione esatta della soluzione è dunque:

Se $\sigma_\theta = 0$:

$$L = \prod_i (1 + \cos^2\theta_i + A\cos\theta_i) \epsilon(\theta_i)$$

$$\Rightarrow -\log L = -\sum_i \log \epsilon_i - \sum_i \log(1 + \cos^2\theta_i + A\cos\theta_i)$$

Sotto queste ipotesi non è necessario conoscere l'efficienza. Trovando il minimo:

$$0 = \frac{\partial -\log L}{\partial A} = - \sum_i \frac{\cos\theta_i}{1 + \cos^2\theta_i + A\cos\theta_i}$$

per $A \ll 1$ (piccola asimmetria):

$$\sum_i \frac{\cos\theta_i}{1 + \cos^2\theta_i} \left(1 + A \frac{\cos\theta_i}{1 + \cos^2\theta_i}\right) = 0$$

$$A = \frac{- \sum_i \frac{\cos\theta_i}{1 + \cos^2\theta_i}}{\sum_i \frac{\cos^2\theta_i}{1 + \cos^2\theta_i}}$$

Introducendo la diluizione D

$$L(A)(1 - D) + DL(-A) = \Pi_i (1 + \cos^2\theta_i + (1 - 2D)A\cos\theta_i)\epsilon_{\theta_i}$$

Esercizio 13

Si consideri di testare l'ipotesi che tutte le misure siano misure della stessa quantità x_0

$$\chi^2 = \sum_i \left(\frac{x_i - x_0}{\sigma_i}\right)^2 \Rightarrow x_0 = \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}$$

$$V(x_0) = \frac{1}{\sum_i \frac{1}{\sigma_i^2}}$$

Combinazioni con alto errore contribuiscono poco, ma danno sempre un miglioramento a meno che non abbiano un grande errore sistematico.

Esercizio 14

Mediando le misure di LEP ed SLD secondo il risultato dell'esercizio 1 si ottiene

$$LEP \quad \sin^2\theta_w = 0.23196 \pm 0.00028$$

$$SLD \quad \sin^2\theta_w = 0.23152 \pm 0.00041$$

La differenza tra le due misure è allora

$$d = 0.00044 \pm 0.00050$$

Tramite il test di Student è possibile testare l'ipotesi che questa quantità sia consistente con un valor medio 0. Si definisce infatti

$$t = (d - \mu) / \sqrt{(s^2 * (N - 1))}$$

dove $\mu = 0$, $N = 2$, $\sqrt{(s^2)} = 0.00050$

$$t = 0.88$$

Bisognerebbe avere una tabella per trovare la probabilità...

Esercizio 15

$$P(0) = e^{-L\sigma\epsilon} > 0.95 \text{ e dunque } \sigma = -\frac{\log 0.95}{\epsilon}$$

Esercizio 16

$$P^{2bin}(N \geq n_{oss}) = \sum_{k=10}^{\infty} P_{\mu=1}(k) = 1 - \sum_{k=0}^9 P_{\mu=1}(k)$$

$$P(N \geq n_{oss})|_{ovunque} = 10(1 - \sum_{k=0}^9 P_{\mu=1}(k)) = 10(1 - \sum_{k=0}^9 \frac{e^{-1}}{k!})$$

Esercizio 17

Facendo uso del teorema della probabilità composta $P(B, A) = P(B|A)P(A)$:

$$\begin{aligned} P_N(\mu_S | \mu_B) &= \frac{P_N(\mu_S + \mu_B)}{P_N(\mu_B)} = \frac{\frac{e^{-(\mu_S + \mu_B)} (\mu_S + \mu_B)^k}{k!}}{\frac{e^{-\mu_B} \mu_B^k}{k!}} \\ &= e^{-\mu_S} (1 + \frac{\mu_B}{\mu_S})^k \end{aligned}$$

La probabilità di vedere meno eventi di quanti se ne osservano (n_0) è:

$$P_{N \leq n_0}(\mu_S | \mu_B) = e^{-\mu_S} \frac{\sum_0^{n_0} \frac{(\mu_S + \mu_B)^k}{k!}}{\sum_0^{n_0} \frac{\mu_B^k}{k!}}$$

Il segnale può essere escluso se $P_{N \leq n_0} < 5\%$.

Nel nostro caso $\mu_S = 4$ e $\mu_B = 1$ e quindi:

$P_{N \leq n_0}$	n_0
0.018	0
0.055	1
0.136	2

Basta osservare un evento per dover rinunciare a mettere un limite. Non c'è sensibilità per il segnale.

Esercizio 18

La sensibilità è la sezione d'urto per la quale il livello di confidenza per l'ipotesi di (S+B) è il 5% nel caso non ci sia segnale.

$$\begin{aligned}
 5\% &= \sum_0^{\infty} e^{-\mu_B} \frac{\mu_B^b}{b!} P_{N \leq b}(\mu_S | \mu_B) \\
 &= \sum_0^{\infty} e^{-\mu_B} \frac{\mu_B^b}{b!} e^{-\mu_S} \frac{\sum_0^b \frac{(\mu_S + \mu_B)^k}{k!}}{\sum_0^b \frac{\mu_B^k}{k!}}
 \end{aligned}$$

Bibliografia

- [1] A. G. Frodesen, O. Skjeggstad and H. Tofte, "Probability And Statistics In Particle Physics," *Bergen, Norway: Universitetsforlaget (1979) 501p.*