

# Introduzione all'Analisi dei Dati

PRIMA LA TEORIA, POI ...

*Belle II* Masterclass ~ 2021 03 17

Giulia Casarosa



Istituto Nazionale di Fisica Nucleare

# Oggi Ricostruirete dei Decadimenti...

... utilizzando i dati veri di *Belle II*

➔ farete la ricostruzione dei seguenti decadimenti:

1  $\pi^0 \rightarrow \gamma\gamma$

2  $\phi \rightarrow K^+K^-$

3  $K_S \rightarrow \pi^+\pi^-$

4  $D^{*+} \rightarrow D^0\pi^+; \quad D^0 \rightarrow K^\pm\pi^\mp$

5  $B^+ \rightarrow J/\psi K^+; \quad J/\psi \rightarrow \mu^+\mu^-, e^+e^-$

➔ oggi pomeriggio vedremo gli strumenti che avrete a disposizione per l'analisi, ora vediamo quali sono le basi dell'analisi dati

# Introduzione

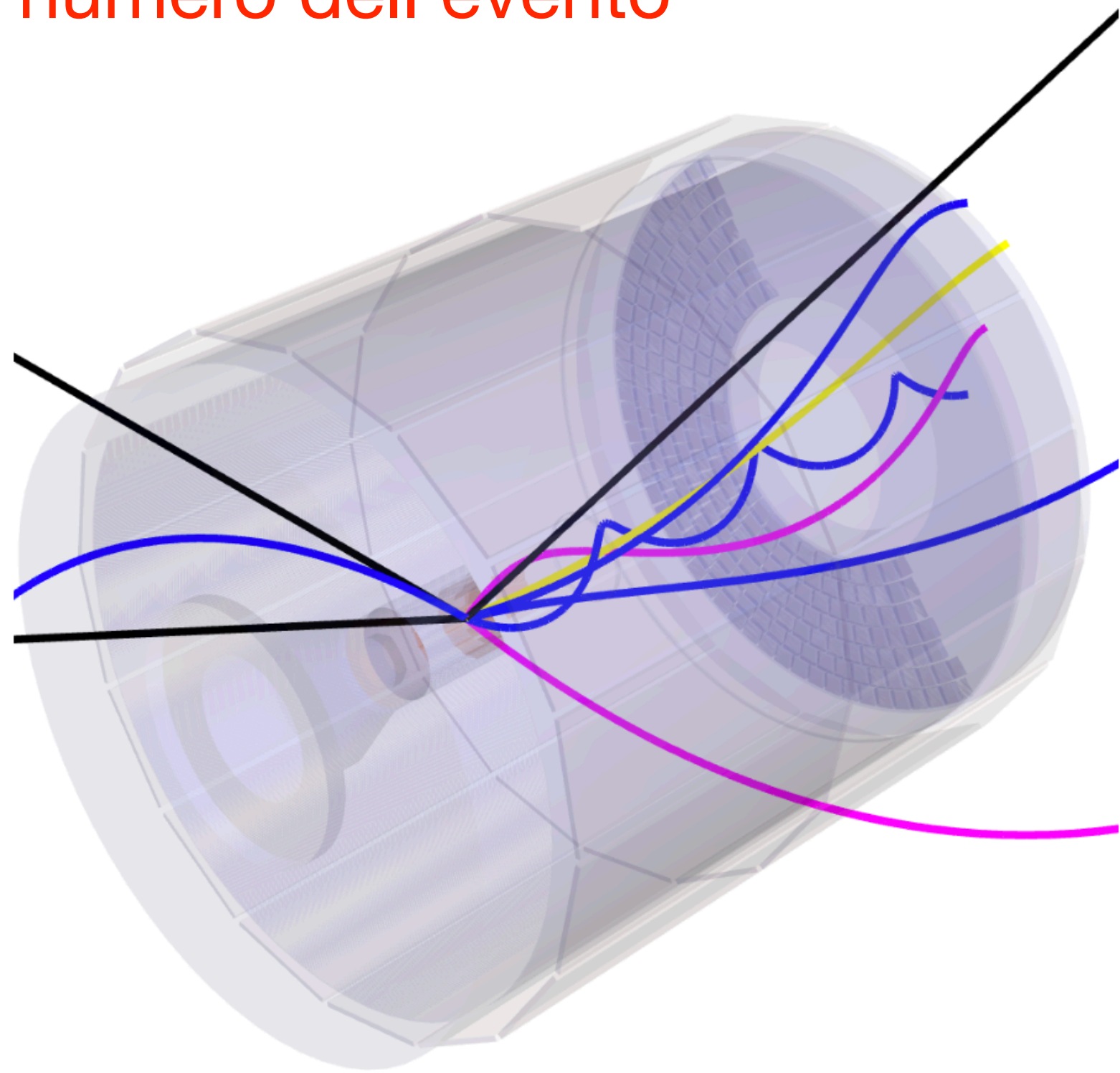
prima di partire...

- ➔ Nelle collisioni  $e^+e^-$  possono accadere tante cose diverse, ci sono **tanti tipi di eventi** che vengono visti dal detector
- ➔ La prima parte dell'analisi dati serve per **selezionare gli eventi interessanti** che vogliamo studiare tra i tantissimi tipi di eventi prodotti nelle collisioni
- ➔ La seconda parte dell'analisi dati ci permette di **estrarre i parametri del modello dai dati** (ad esempio la massa invariante o la vita media di una particella) per poi confrontarli con il modello teorico (metodo scientifico):
  - se i valori che misuro sono compatibili con il modello allora non ho nessun motivo per dire che il modello è sbagliato (e me lo tengo)
  - se i valori che misuro non sono compatibili con il modello, allora devo modificare il mio modello o sceglierne un altro che descriva le osservazioni sperimentali



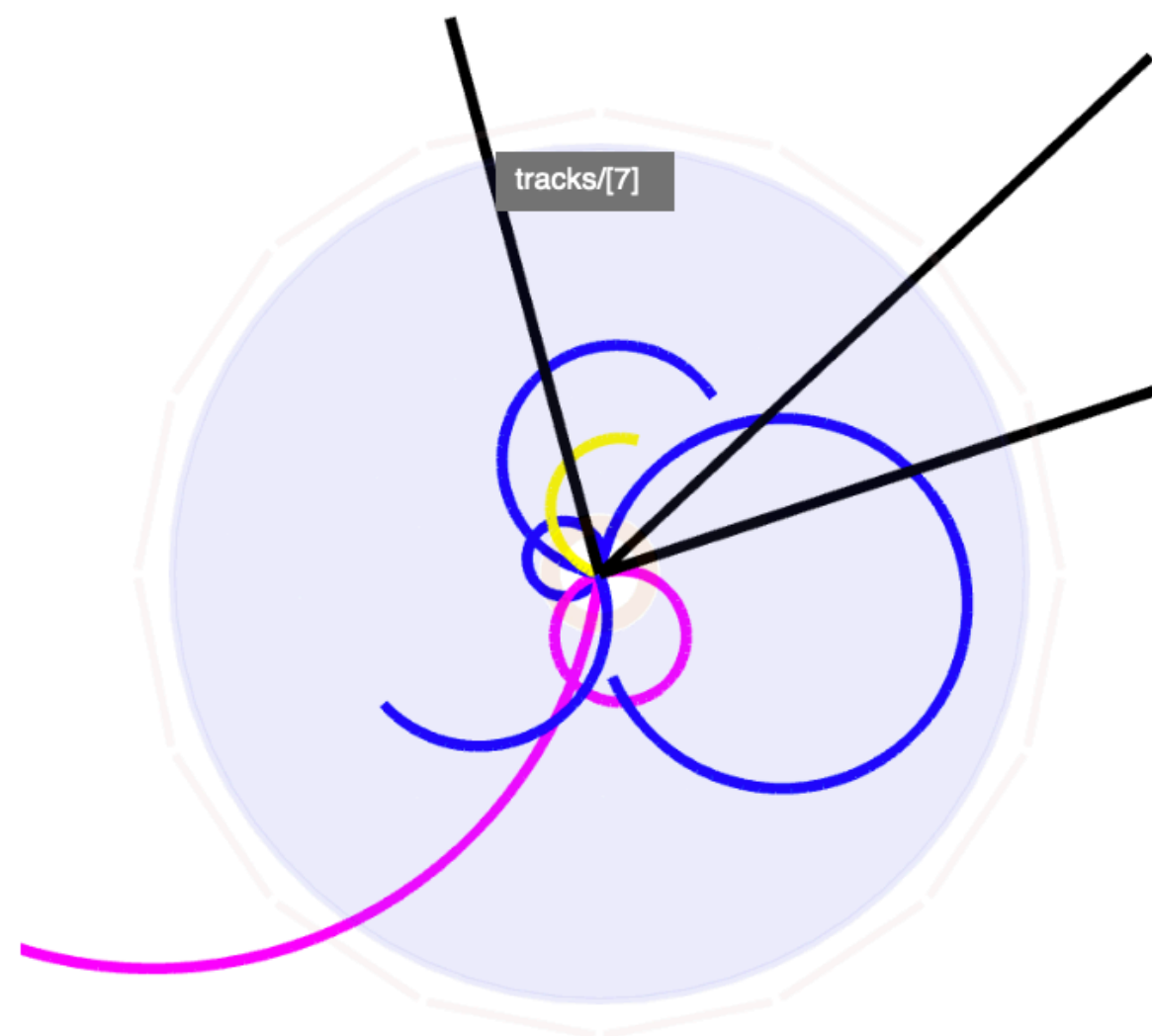
Event: < 8 > Close Window

numero dell'evento

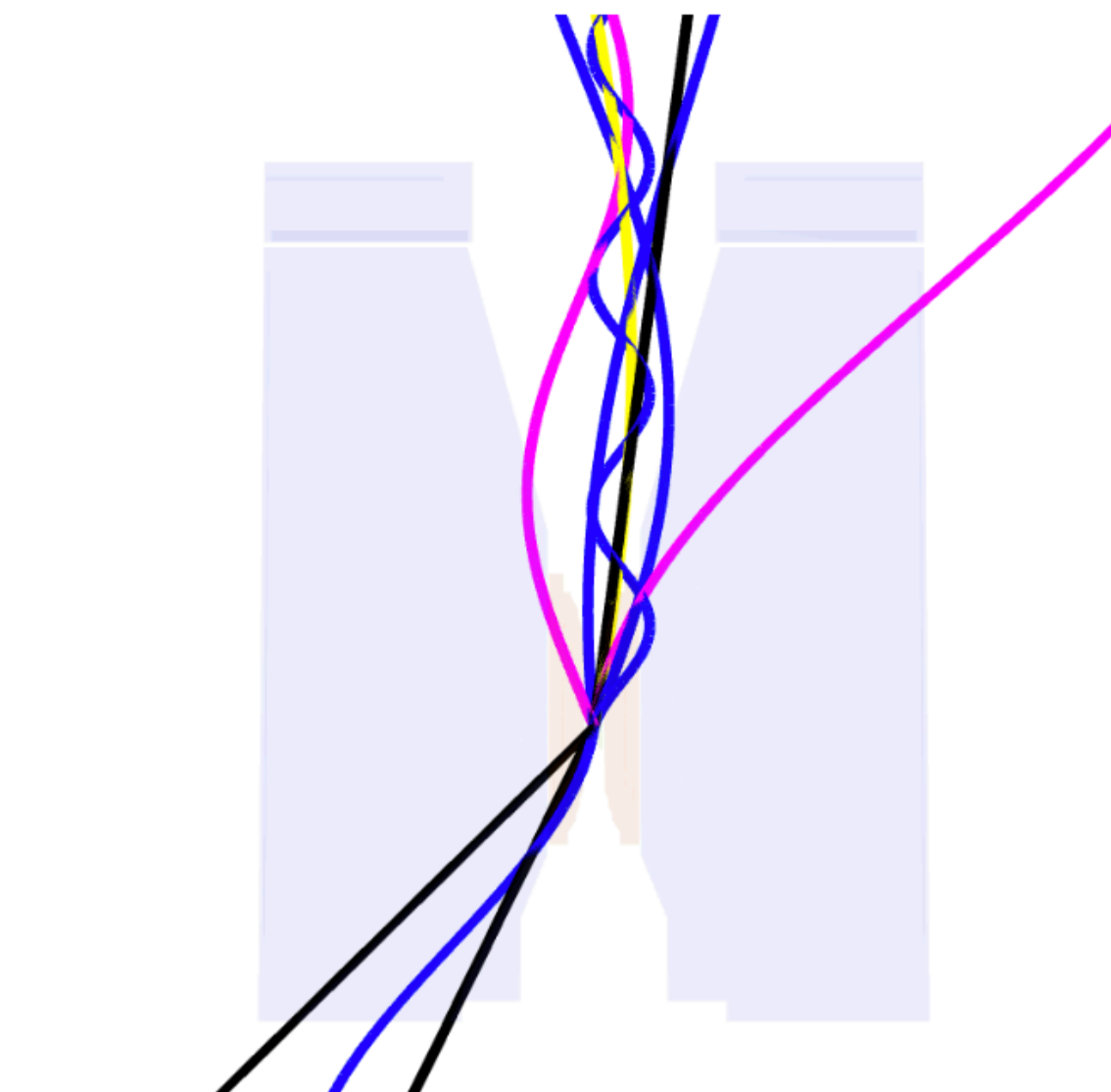


vista 3D del detector e delle particelle finali, zoommabile e girabile col mouse

vista sul piano trasverso



vista sul piano longitudinale



lista delle particelle finali:

Reconstructed particles of Event 8

N	px(GeV/c)	py(GeV/c)	pz(GeV/c)	p(GeV/c)	Energy(GeV)	ChargeID	
1	-0.0844734	-0.194447	0.0572224	0.21959	0.260192	-1	pion
2	0.263173	0.0881825	0.189857	0.336276	0.996713	1	proton
3	-0.0845184	-0.683124	0.468224	0.832488	1.25435	1	proton
4	-0.354341	0.0548067	0.345443	0.497887	0.51708	-1	pion
5	-0.259805	0.0805835	0.445737	0.522182	0.718604	1	kaon
6	0.0741679	0.45684	-0.236536	0.519762	0.538175	1	pion
7	0.124356	-0.341834	0.502423	0.620277	0.635786	-1	pion
8	-0.0309192	0.11498	0.234523	0.263016	0.263016	0	photon
9	0.36041	0.338427	-0.354005	0.608069	0.608069	0	photon
100	0.118693	0.0391602	-0.240812	0.271315	0.271315	0	photon

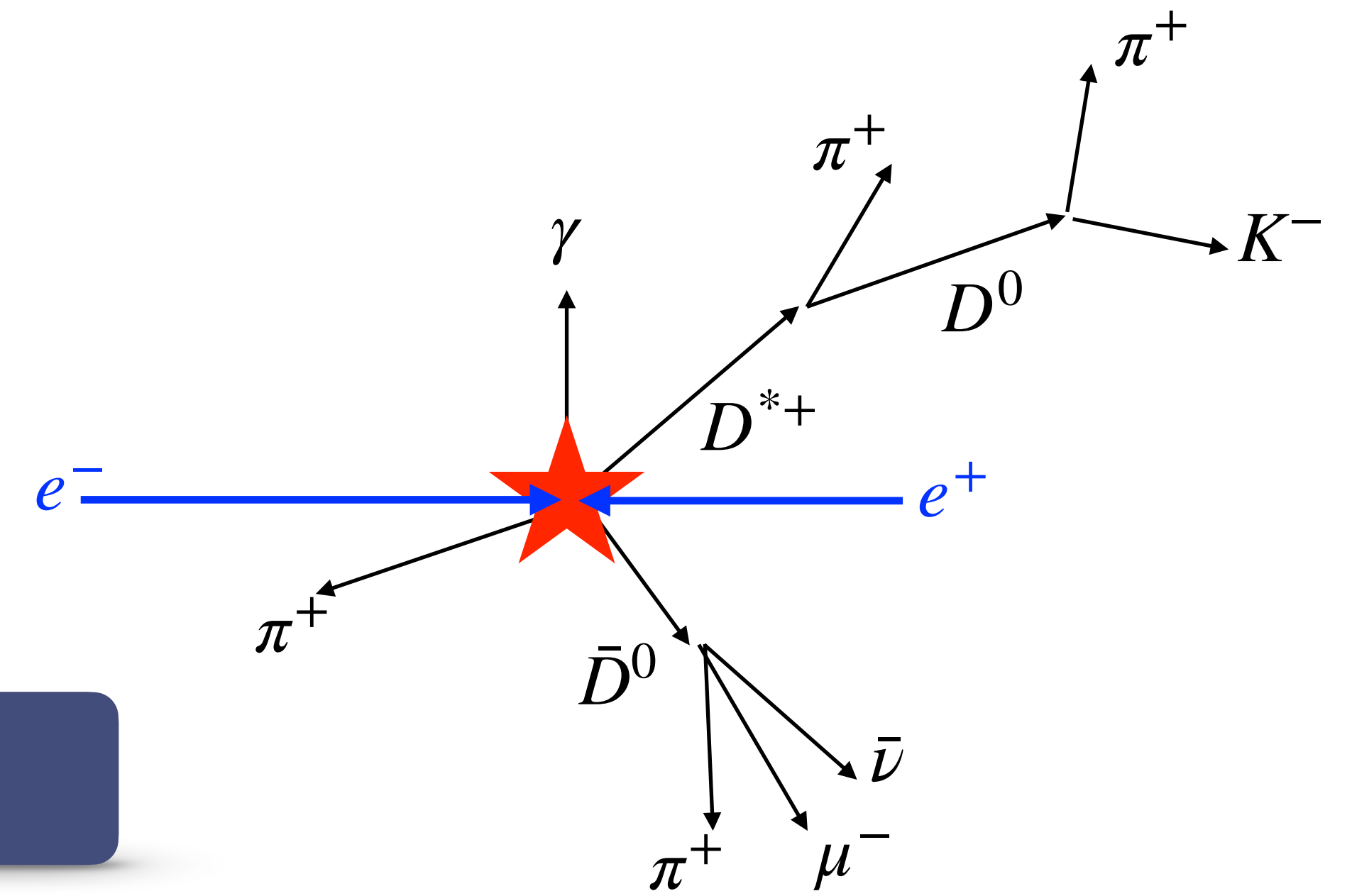
di queste particelle conosciamo:

1. direzione
2. quantità di moto
3. energia
4. carica
5. tipo

vediamo qualche esempio (interattivo) di evento visto con l'Event Display

# Analisi Dati

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$



particelle finali

ricostruzione

decadimento completo

selezione

candidati interessanti

istogramma

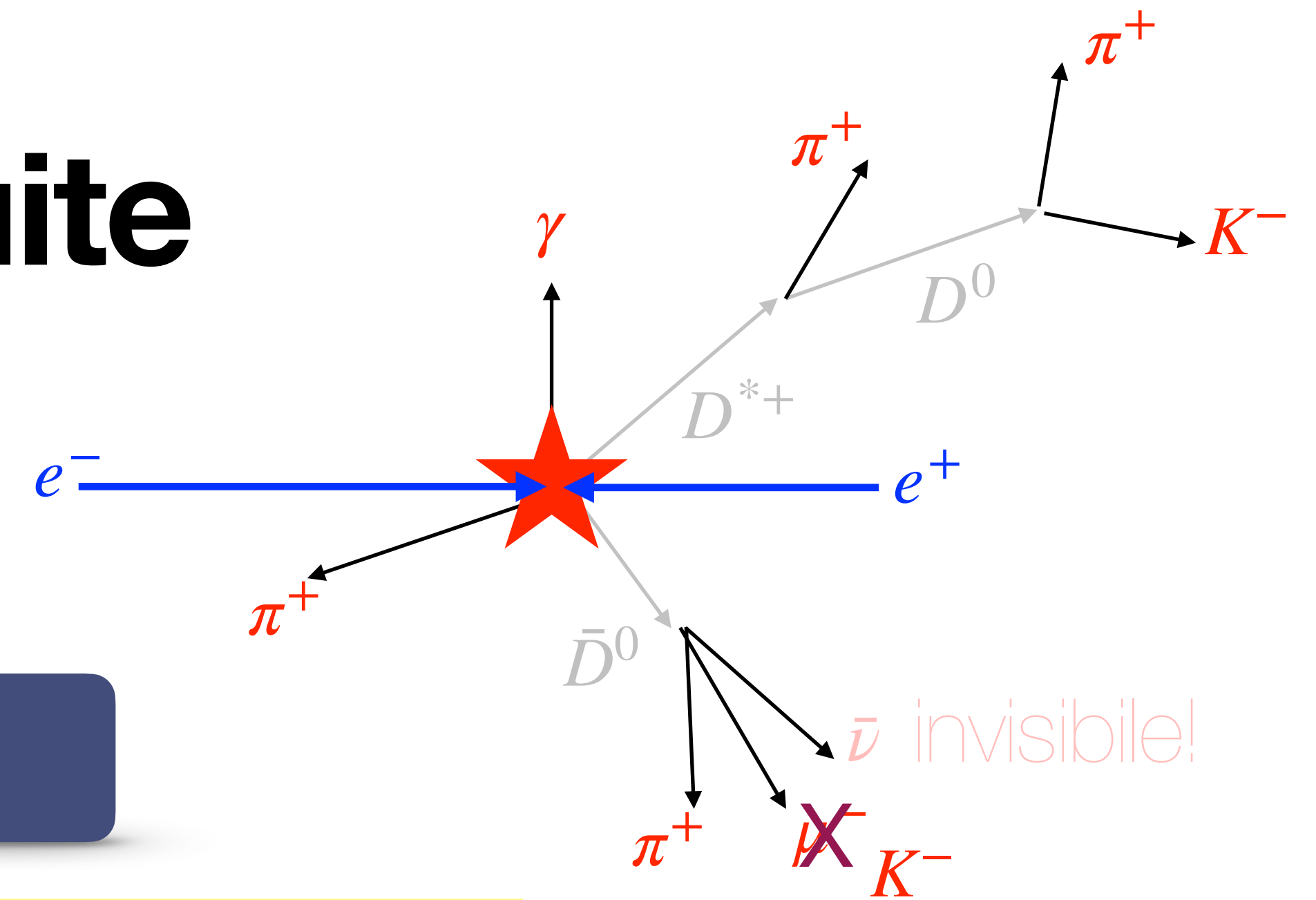
fit

osservabili misurate

modello

# Partenza: particelle ricostruite

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$



particelle finali

ricostruzione

decadimento completo

## → particelle finali

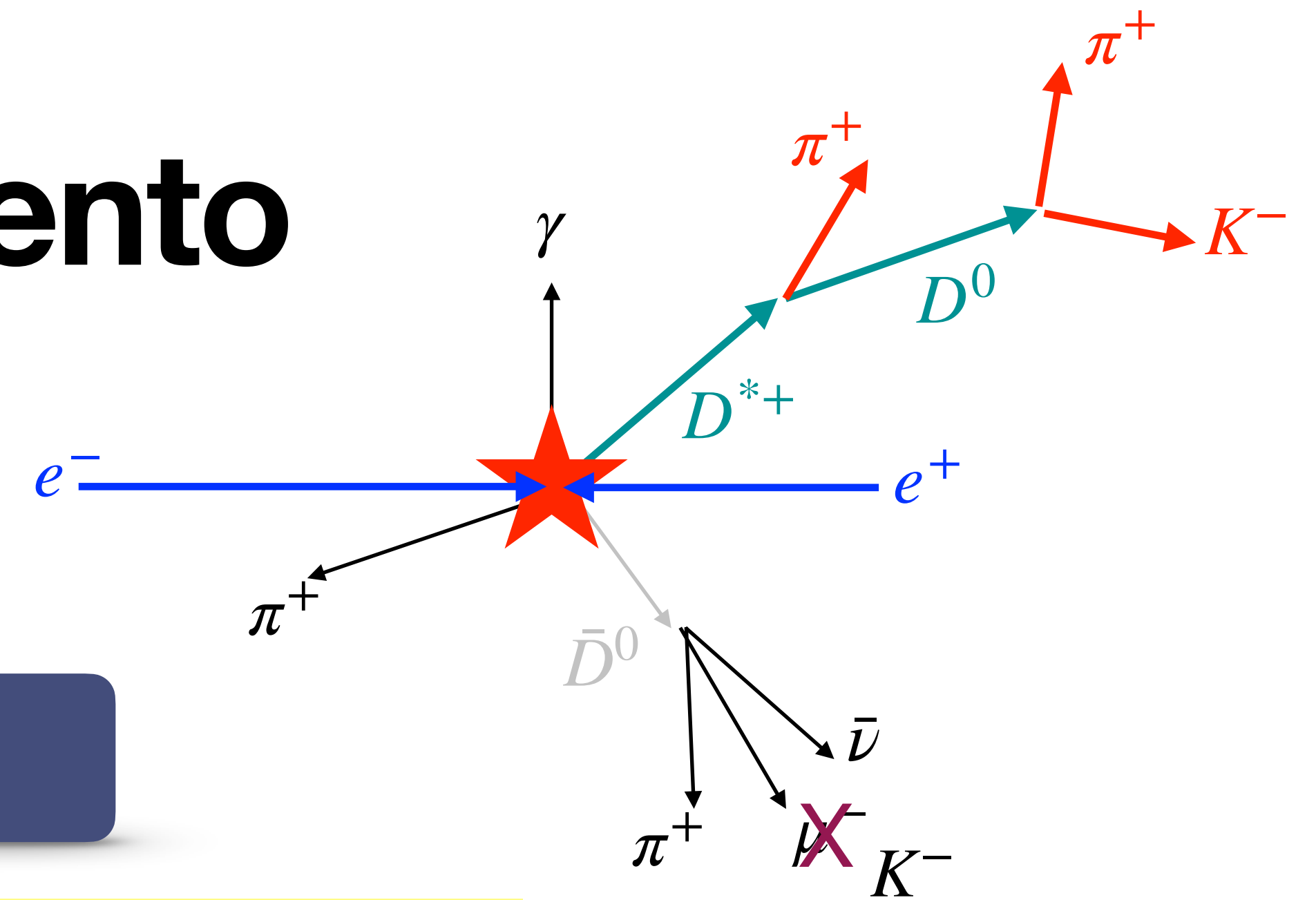
- quelle particelle che hanno attraversato il detector lasciando un segnale del loro passaggio
- la ricostruzione non è perfetta, è possibile:
  - perdere delle particelle, non ricostruirle
  - sbagliare il tipo di particella, ad esempio  $\mu^- \rightarrow K^-$
  - sbagliare l'energia, la direzione, ... delle particelle
- alcune particelle sono invisibili! ad esempio il  $\nu$

modello



# Ricostruzione del Decadimento

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$



particelle finali

ricostruzione

decadimento completo

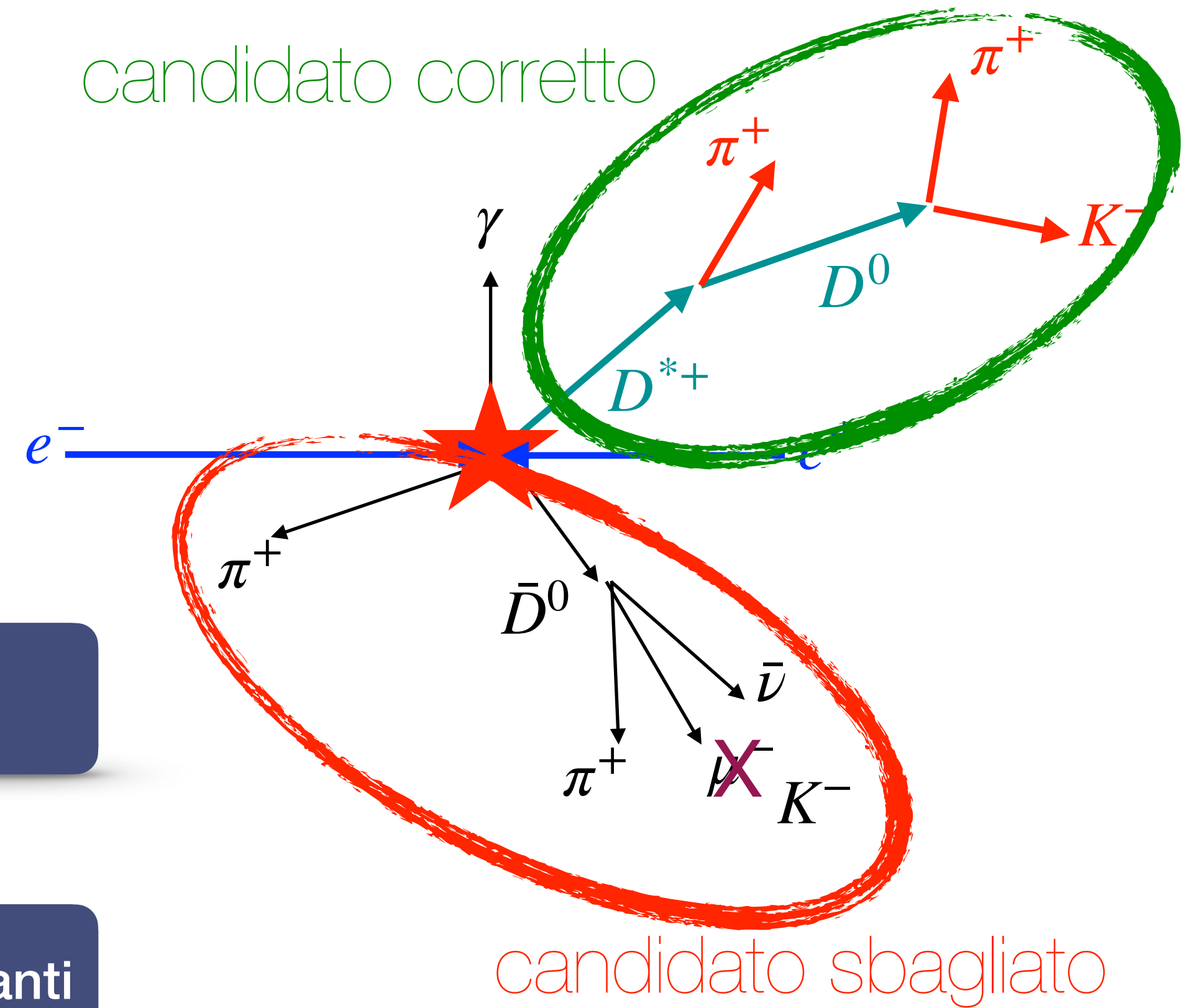
## → ricostruzione del decadimento di interesse

- si combinano le particelle finali per ottenere la catena di decadimento
  - si combinano prima tutte le coppie di  $\pi^+$  e  $K^-$  alla ricerca di **candidato**  $D^0$
  - poi si combinano i candidati  $D^0$  con i  $\pi^+$  avanzati, alla ricerca di un candidato  $D^{*+}$
- molti di questi candidati non saranno corretti!

modello

# Selezione dei Candidati

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$



particelle finali

ricostruzione

decadimento completo

selezione

candidati interessanti

## ➔ selezione dei candidati

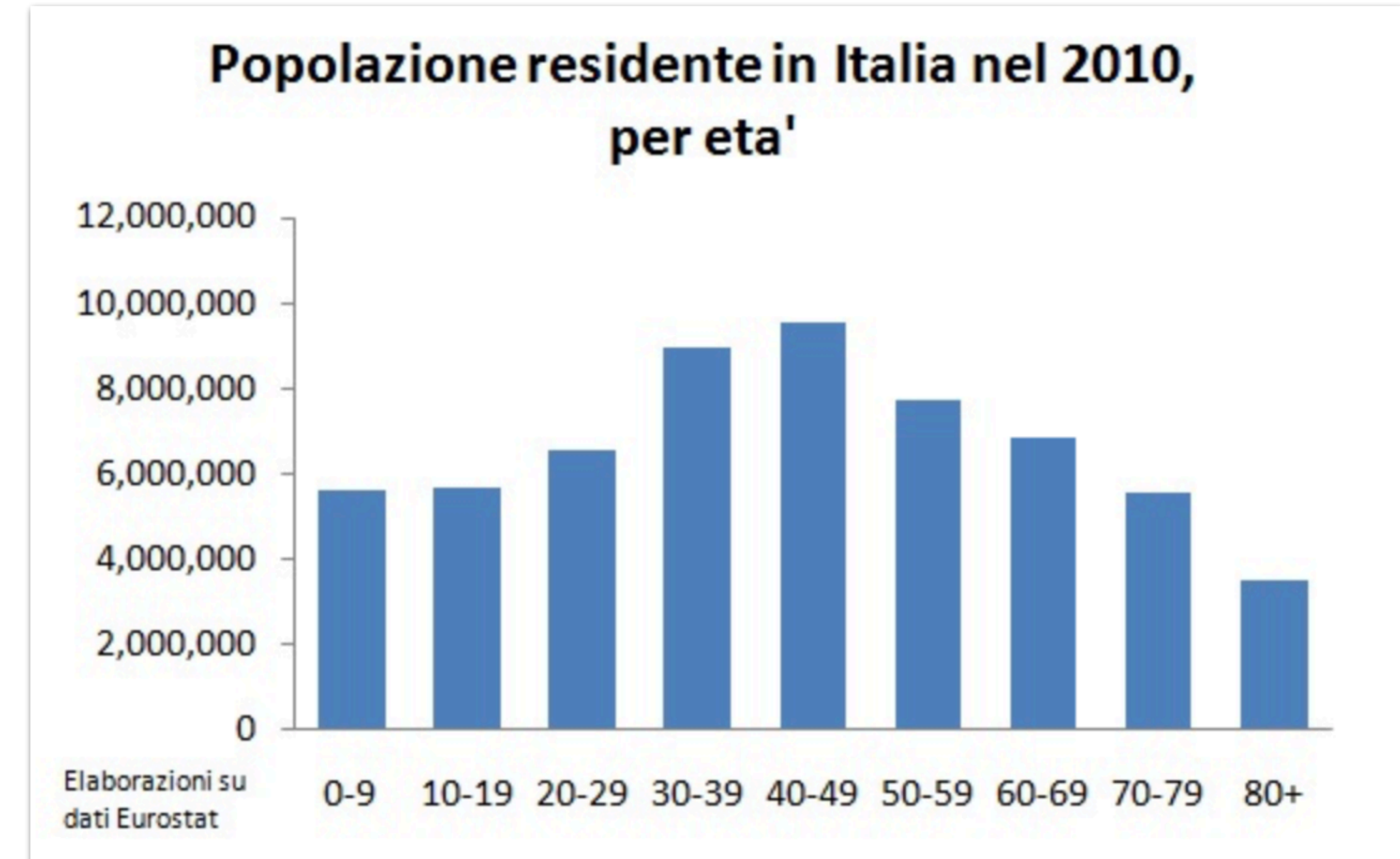
- obiettivo: rimuovere la maggior parte dei candidati “sbagliati” e tenere la maggior parte dei candidati corretti
- consiste nel richiedere che le particelle (finali o ricostruite) abbiano certe caratteristiche
  - carica, massa invariante dei candidati ricostruiti ( $D^0$ ,  $D^{*+}$ ), quantità di moto/energia, ...



# L'istogramma

uno strumento estremamente utile per “guardare” i dati

- ➔ gli eventi non si possono guardare uno per uno!
- ➔ l'istogramma è uno strumento che ci permette di vedere tutti i dati insieme
  - l'istogramma è una rappresentazione dei dati
- ➔ ad esempio, guardiamo l'istogramma dell'età della popolazione residente in Italia nel 2010
  - ogni colonna (bin) ci dice quanti italiani hanno un'età compresa tra 0-9, 10-19, ... anni
  - l'istogramma è caratterizzato da un minimo e un massimo dell'asse x, e dal numero di bin



# Candidati Interessanti

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$

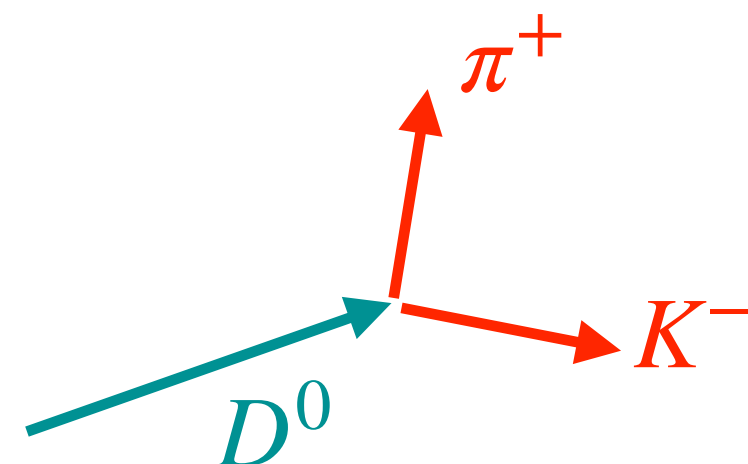
→ come faccio a sapere se sto selezionando i candidati corretti?

- ci sono diversi metodi, nella maggioranza delle analisi si guarda la distribuzione della massa invariante (istogramma)

$$M_{K\pi} c^2 = \sqrt{(E_K + E_\pi)^2 - |\vec{p}_K + \vec{p}_\pi|^2 c^2}$$

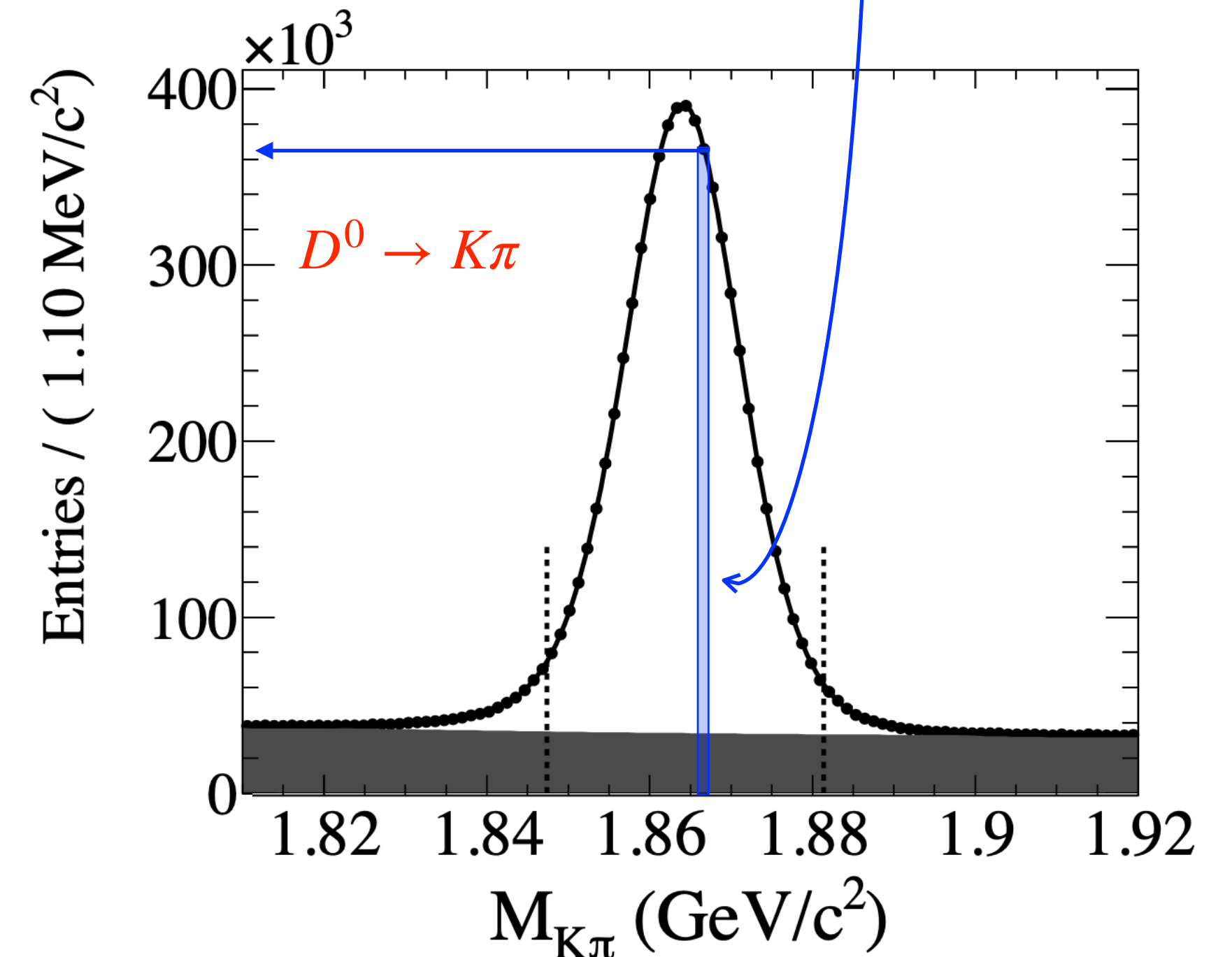
massa  
invariante  
del  $D^0$

energia & quantità di moto  
delle particelle figlie,  $\pi^+ K^-$



ci sono  $\sim 360 \times 10^3$  candidati  $D^0$  con una massa invariante nel bin dell'istogramma

distribuzione della massa invariante del  $D^0$ :



negli istogrammi è importante controllare:

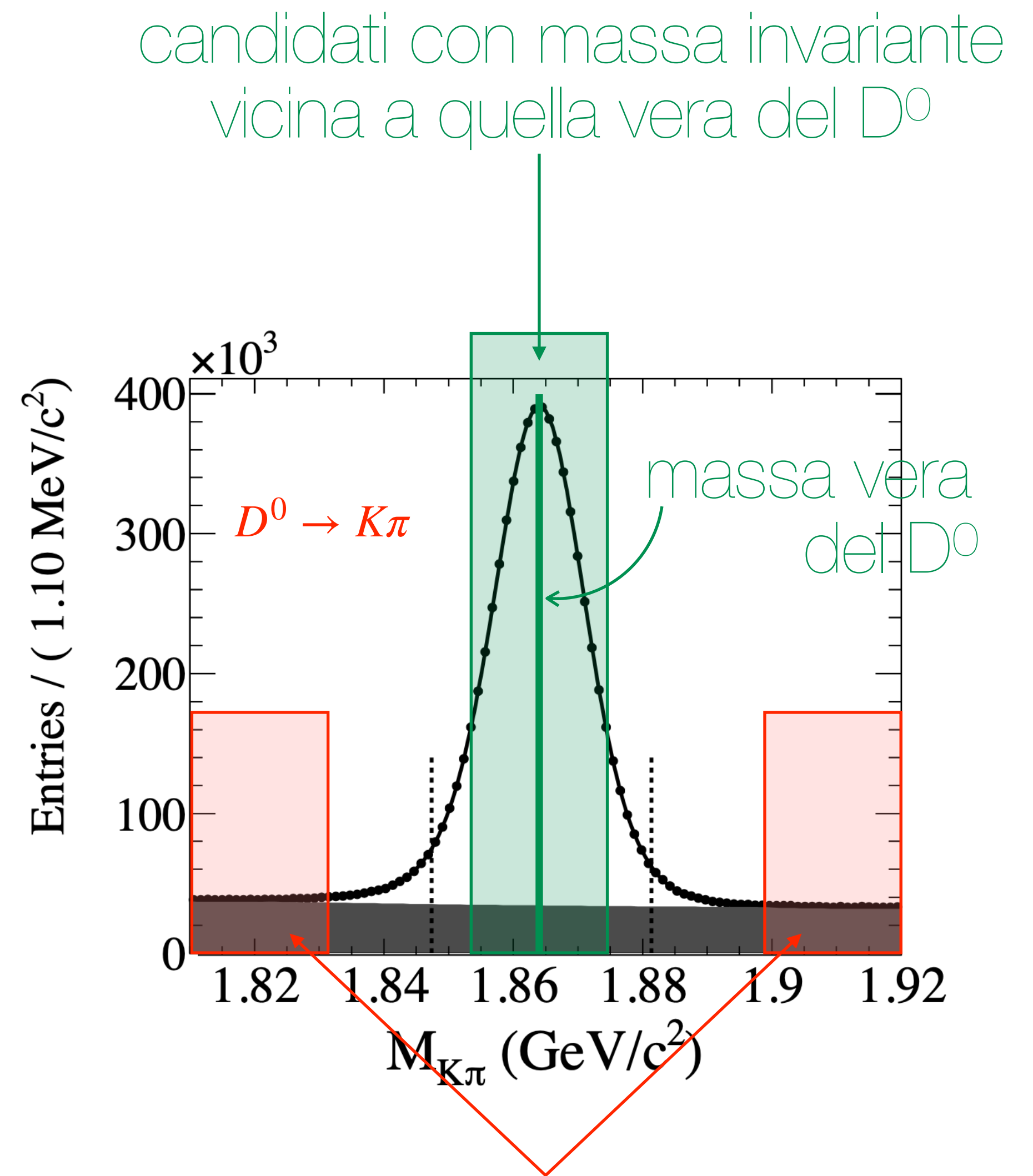
- il minimo e il massimo (i miei candidati ci sono tutti o alcuni non li vedo perchè stanno fuori dal range?)
- il numero di bin utilizzati (nella discretizzazione mi posso perdere delle informazioni, oppure posso discretizzare troppo e non vedere niente)

# Candidati Interessanti

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$

➔ come faccio a sapere se sto selezionando i candidati corretti?

- ci sono diversi metodi, nella maggioranza delle analisi si guarda la distribuzione della massa invariante (istogramma)
- se il candidato è un  $D^0$  vero allora la massa invariante del candidato sarà vicina alla massa vera del  $D^0$
- perchè i candidati correttamente ricostruiti non hanno un valore di massa invariante *uguale* a quella vera?

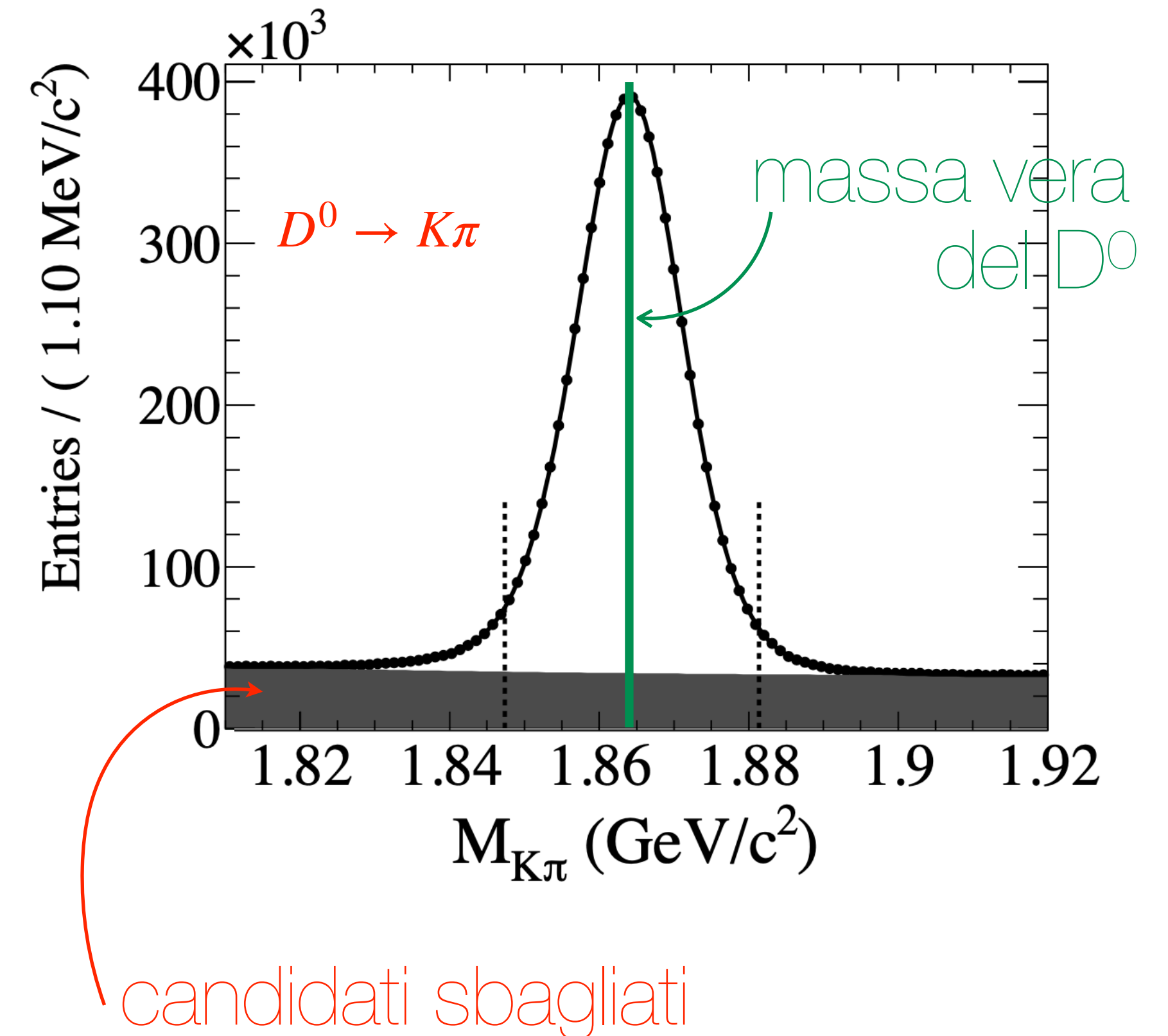




# Il Background

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$

- ➔ alcuni candidati che passano la selezione non sono corretti
- ad esempio, nella lista di candidati dopo la selezione è possibile che alcuni candidati abbiano la massa invariante giusta *per caso*
- per distinguere il background dal segnale si possono guardare distribuzioni di altre variabili
- il nostro obiettivo è estrarre i parametri dai dati corrispondenti ai candidati corretti, come si fa con quelli sbagliati?
- nel modello di fit che usiamo per estrarre i parametri dai dati dobbiamo tener conto della presenza di questi candidati sbagliati

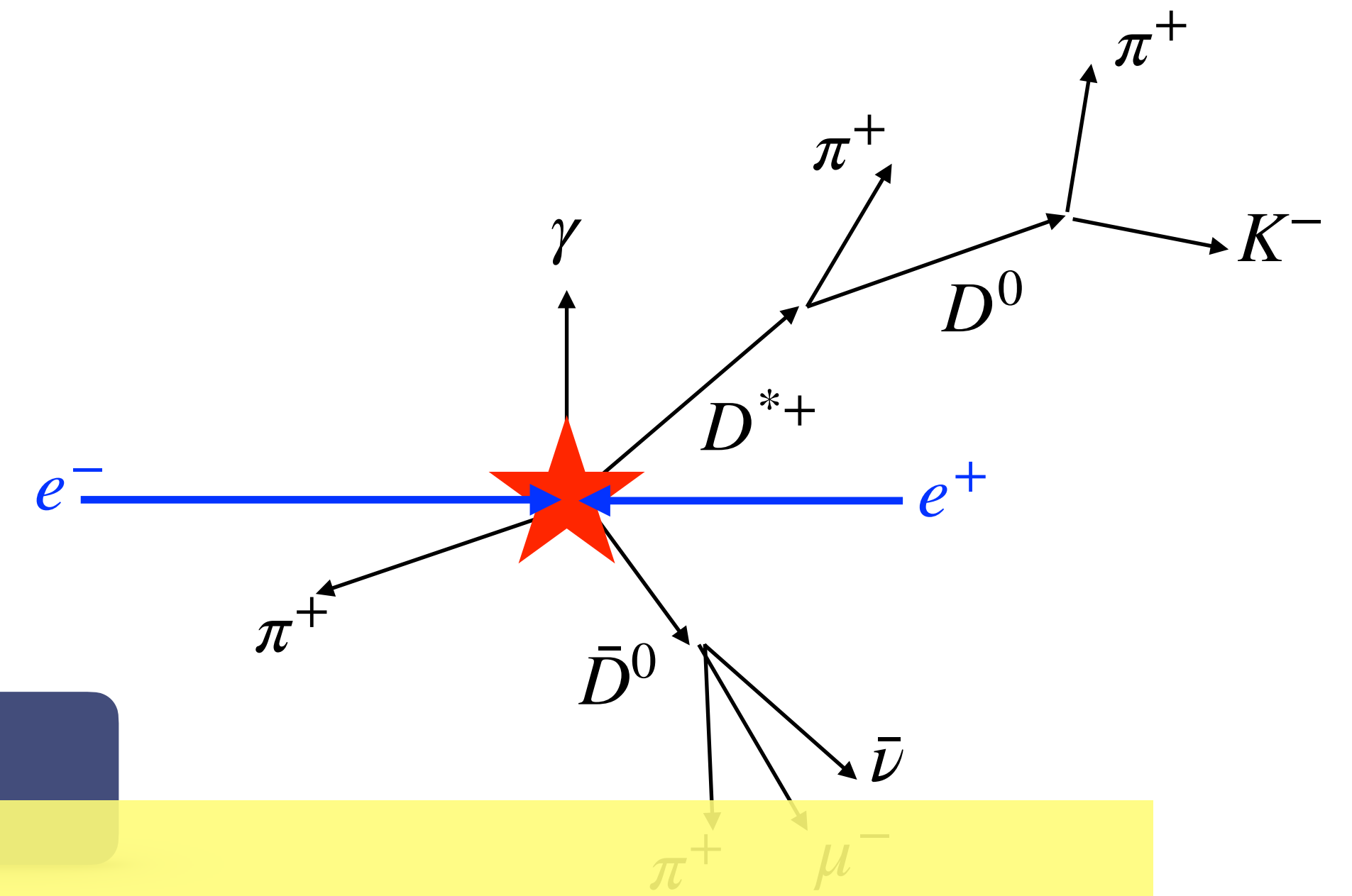


l'analisi statistica è una parte molto importante dell'analisi dati



# Analisi Dati

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$



particelle finali

ricostruzione

decadimento completo

## → estrazione delle informazioni interessanti

- questo step dipende dall'analisi
- spesso si fa un fit ai dati assumendo un modello teorico ed estraendo i parametri del modello che *meglio* descrivono i dati ricostruiti
  - i parametri estratti si confrontano con i valori aspettati dalla teoria

istogramma

fit

osservabili misurate

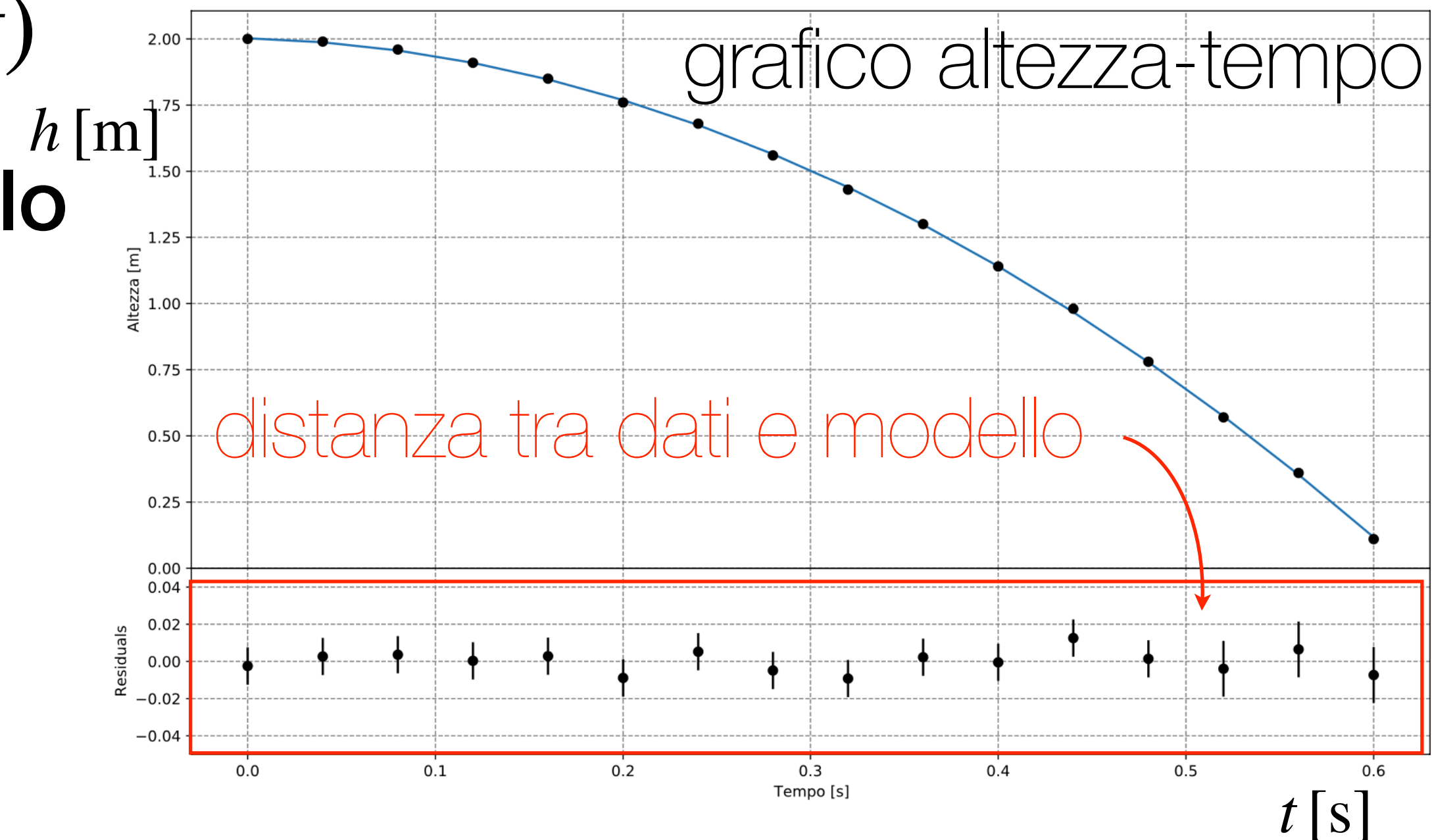
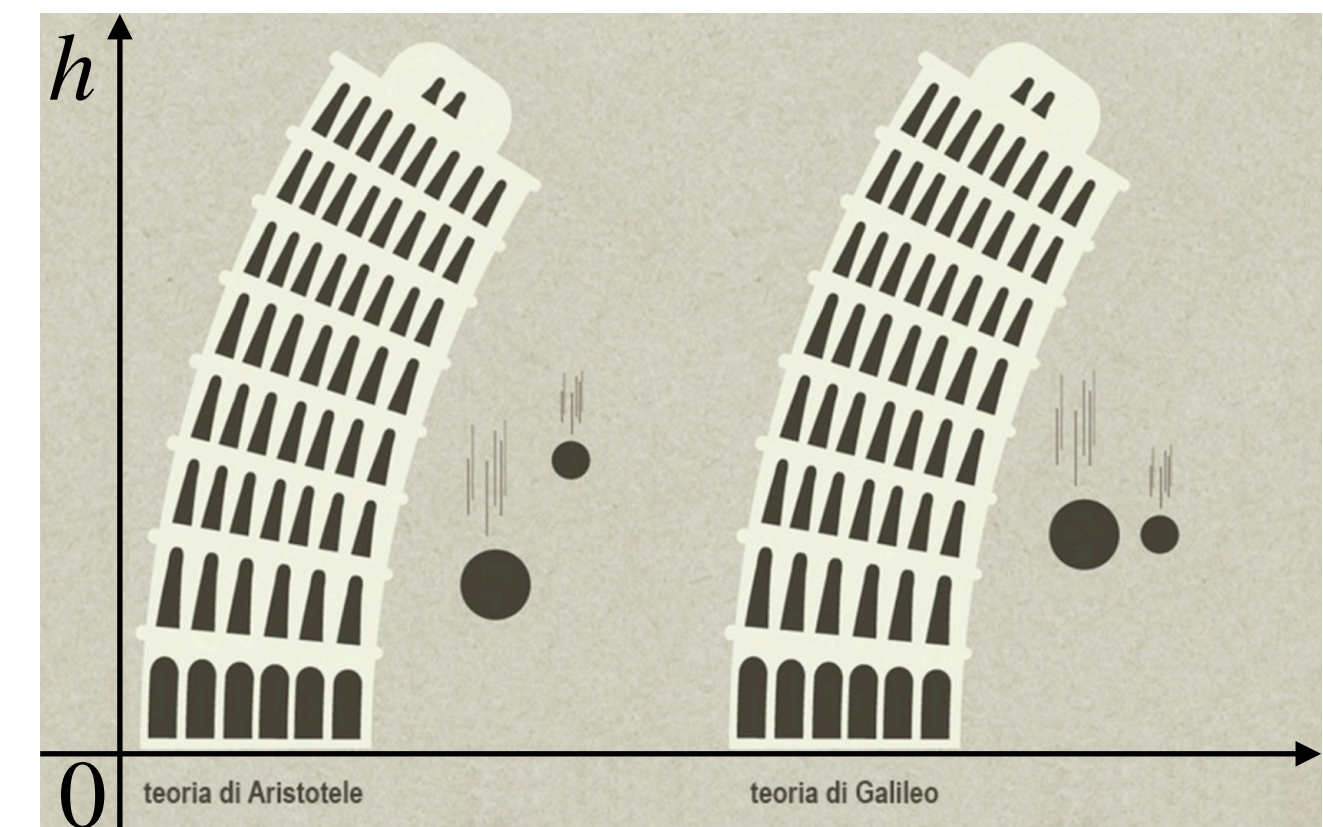
modello

# L'idea del Fit ai Dati

esempio: caduta di un grave

- ➔ Misuro l'altezza  $h$  di un grave che casca a tempi successivi e costruisco il grafico  $h(t)$
- ➔ voglio confrontare queste misure con quello che mi dice la teoria:
  - teoria:  $h(t) = h_0 + v_0 t + \frac{1}{2} a t^2$  (1)
  - mi aspetto:  $h_0 = 2 \text{ m}$ ;  $v_0 = 0$ ;  $a = -9.81 \text{ m/s}^2$
- ➔ faccio il fit: estraggo i parametri di (1) che meglio descrivono i dati sperimentali, trovo:

$$h_0 = (2.002 \pm 0.004) \text{ m}; \quad a = (-9.9 \pm 0.1) \text{ m/s}^2$$



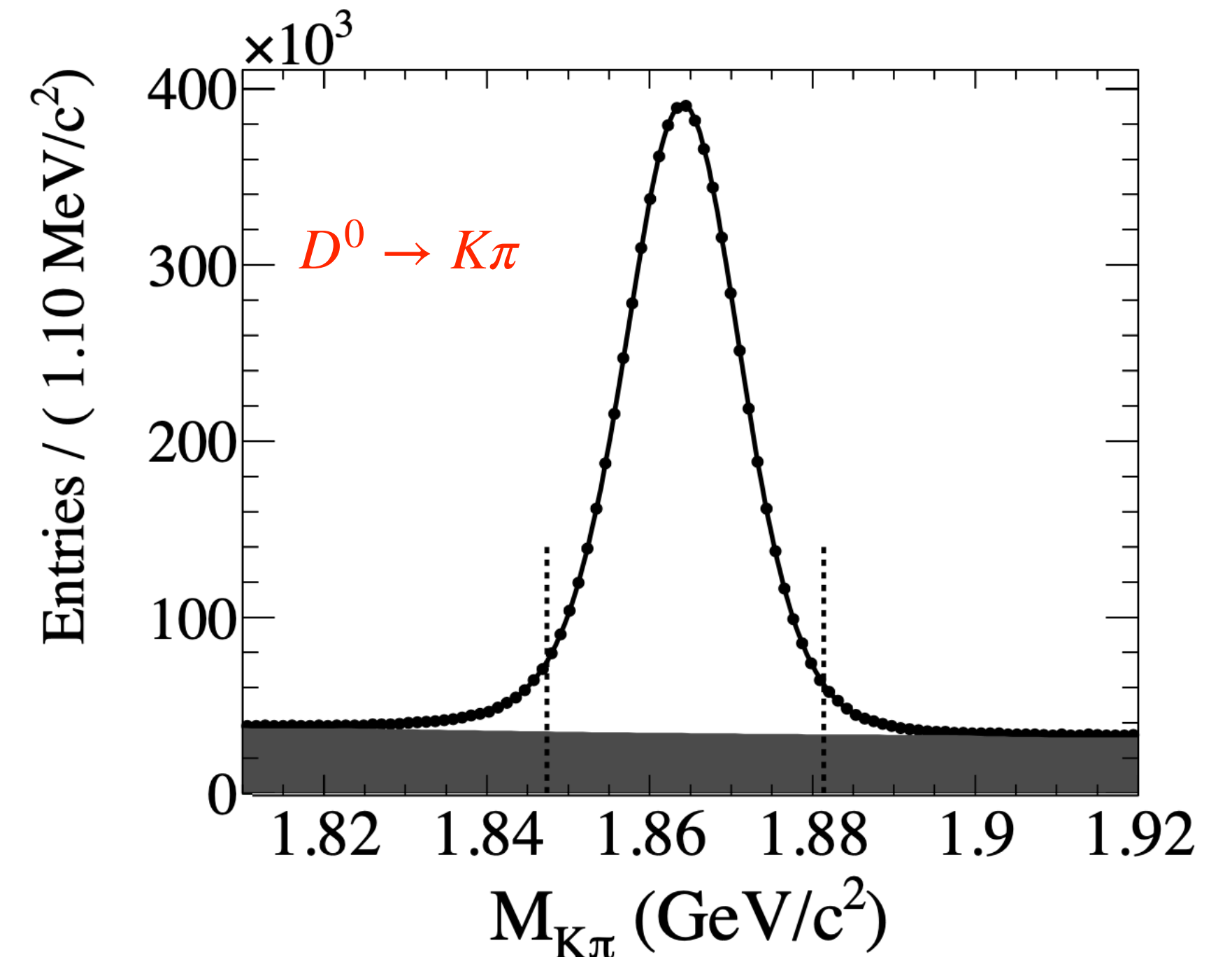
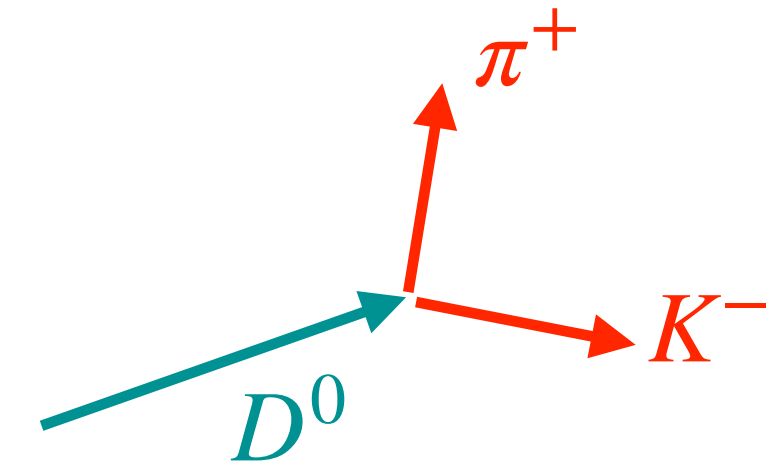


# Fit ai Dati

esempio:  $D^{*+} \rightarrow D^0 \pi^+$ ;  $D^0 \rightarrow K^- \pi^+$

➔ con lo stesso principio posso estrarre informazioni dalle distribuzioni dei candidati ricostruiti

- variabili che si possono fittare: massa invariante, distribuzione angolare, tempo proprio, ...
- esempio, massa invariante del  $D^0$ 
  - calcolo la massa dei candidati  $D^0$  ricostruiti e li metto in un istogramma
  - faccio un fit alla distribuzione in massa invariante per estrarre la massa ( $M_{D^0}$ ) particella
  - le funzione di fit possono essere complicate, con tanti parametri “in più” che servono per descrivere l’effetto del detector ( $\vec{a}_{\text{sig}}$ ) e per descrivere il background ( $\vec{a}_{\text{bkg}}$ )



$$f_{\text{fit}}(m) = \underbrace{N_{\text{sig}} f_{\text{sig}}(m | M_{D^0}, \vec{a}_{\text{sig}})}_{\text{descrizione dei candidati corretti}} + \underbrace{N_{\text{bkg}} f_{\text{bkg}}(m | \vec{a}_{\text{bkg}})}_{\text{descrizione dei candidati sbagliati}}$$



# Ora ci Facciamo un Giro nella **Realtà Virtuale**

