

# How to Reconstruct the Unknown Physical Quantities Using the Bayesian Approach and Neural Networks

*Cracow Computational Science Workshop,  
3-4 April 2006*

**Marcin Wolter**

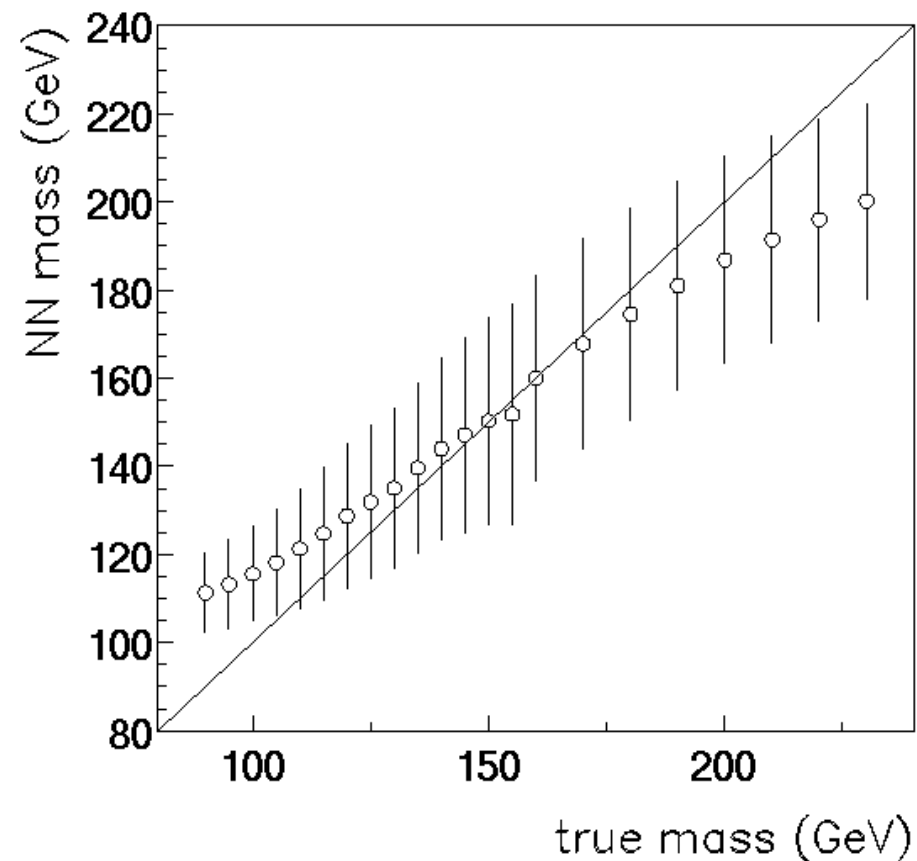
*Institute of Nuclear Physics PAN, Kraków*

# Task

---

- Find the unknown particle mass with no explicit knowledge of the functional dependence of the mass estimate on the measured quantities (but having simulated data).
- **Shortly:**
  - using non-calibrated jet energies and an angle between them,
  - without using the invariant mass formula.
- *Make the optimal use of the information supplied.*
- The example presented shows the mass reconstruction of the hypothetical Higgs boson in the process:  $H \rightarrow b \bar{b} \rightarrow 2 \text{ jets}$

# Let's train a Neural Net to return Higgs mass directly



Feed-forward network, logistic activation function, two hidden layers with 10 nodes in each.

- NN trained on MC samples ( $M_H=90\text{GeV}$  up to  $M_H=230\text{ GeV}$ ). It returns Higgs mass directly.  
**Input:  $E_1, E_2, \cos(\theta_{12})$ , output:  $M_H$**
- NN minimizes  $\chi^2$ , therefore reconstructed masses are shifted towards the middle of the spectrum.
- RMS from 10 GeV up to 23 GeV
- Could become linear if MC generated in a wider Higgs mass range.
- **We need a better method**

# Another approach

---

- Find the probability distributions of observables for various Higgs masses:  $P(x/M_H)$  where  $x=\{E_1, E_2, \cos(\theta_{12})\}$
- Calculate the Higgs mass probability distribution for each event and later for a sample of events.
- Using this approach it is possible to find correct Higgs mass for non-calibrated jet energies.
- No background present (neither physical nor combinatorial).

# Method description

- Generate MC samples for different Higgs masses  
(95 GeV - 155 GeV, 5 GeV increments).
  - Find a  $E_1, E_2$  and  $\cos(\theta)$  3-dimensional distribution for each Higgs mass.
  - Fit the distributions using neural network and normalize them  
-  $P(x|M_H)$
  - For every event find  $P(M_H/x)$  from the Bayes theorem:  
$$P(M_H/x) \propto P(x|M_H) \times P(M_H).$$
  - For  $n$  events: 
$$P_n(M_H | x_1, \dots, x_n) \propto \left( \prod_{i=1}^n P(x_i | M_H) \right) \times P(M_H)$$
- Mean of the  $P_n(M_H/x)$  distribution is chosen as a Higgs mass estimate.
- We assume  $P(M_H)$  is flat over  $M_H$ .

# Fitting probability distributions with a neural network\*

\*L.Garrido, A.Juste, *Comp. Phys. Com.* 115 (1998)

- Add a reference sample of events with flat distribution  $P_{ref}(x)$  in  $E_1, E_2$  and  $\cos(\theta)$ .
- Use the reference sample as a "background" and MC events as a "signal" for NN training (feed-forward, 50 nodes in a hidden layer).
- Since NN estimates the a-posteriori Bayesian probability to belong to class "signal":

$$NN_{out}(x) = P(sig|x) = \frac{a_{sig} P(x|sig)}{a_{sig} P(x|sig) + a_{ref} P(x|ref)}$$

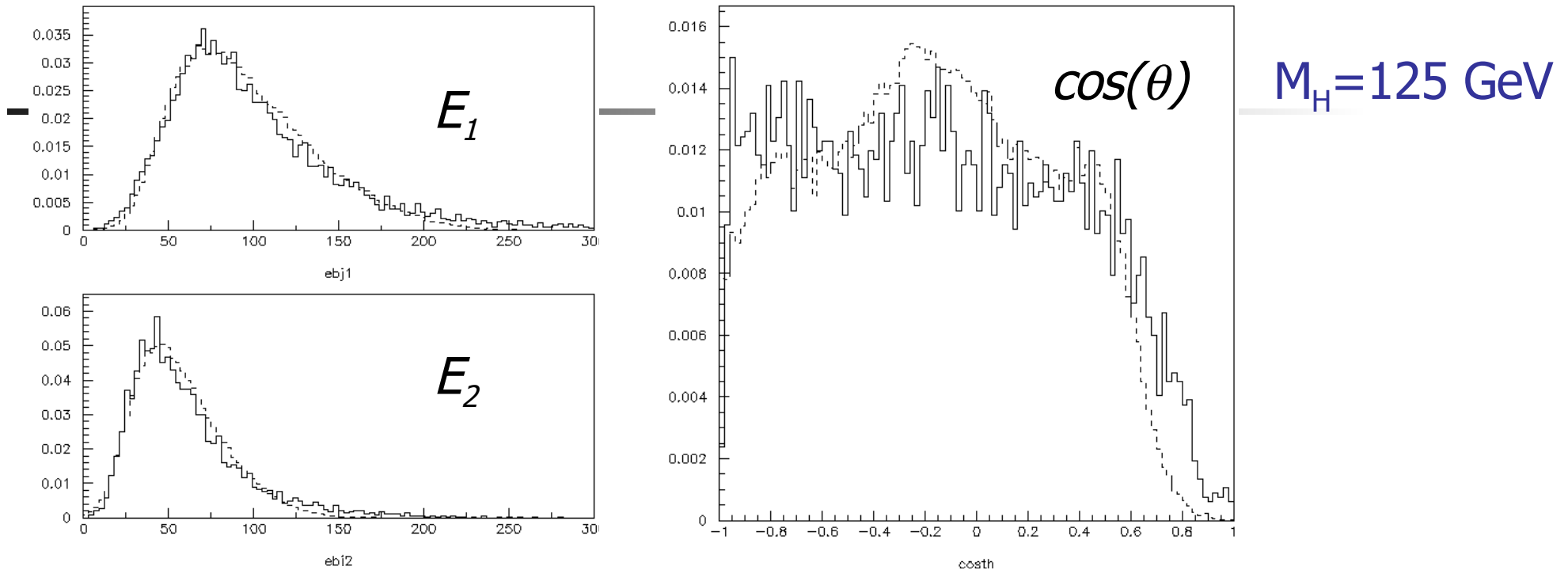
- therefore :

$$P(x|sig) = P(ref|x) \frac{a_{ref}}{a_{sig}} \frac{NN_{out}(x)}{1 - NN_{out}(x)}$$

- where  $a_{ref}$  and  $a_{sig}$  are the proportions of events used for training

$$a_{ref} + a_{sig} = 1$$

# Fit quality



Distributions of  $E_1$ ,  $E_2$  and  $\cos(\theta)$  for  $M_H=125$  GeV sample (solid line) compared with the distributions obtained from the fitted function.

Fit quality could be improved by generating bigger MC samples - here only 2300 events used for a 3D fit. Also another NN activation function (Radial Base?) might help.

**The fit is performed in 3D, not on the projections.**

# Comments on fitting

---

**Advantages of the neural network fit:** unbinned fit, no analytical function needed, trained NN easy to convert into a fast subprogram.

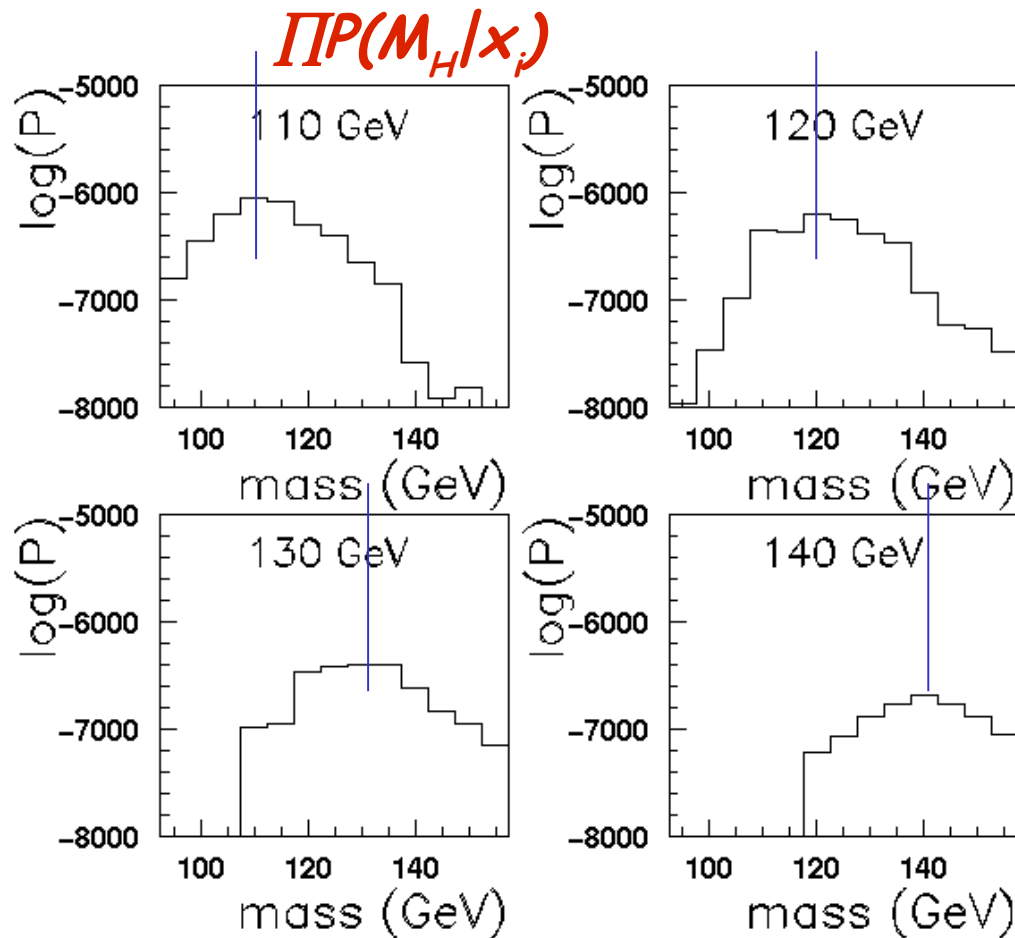
**Alternative methods to fit the distribution:**

histogramming - not enough data/memory when many dimensions.

fit using analytical function - problems finding a suitable function.

kernel-based probability density function - worthwhile to try.

# Higgs mass probability distributions



- For a given event each distribution gives  $P(x|M_H)$ :

$$P(M_H|x) = P(x|M_H) * P(M_H)$$

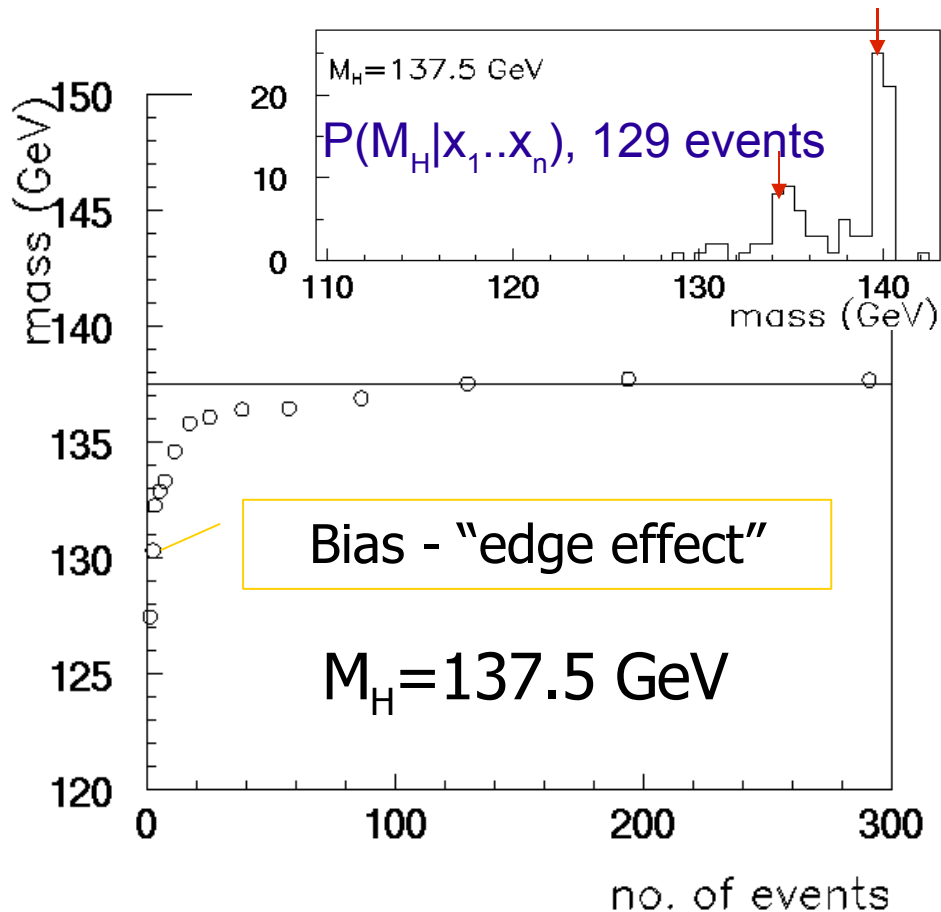
- For a set of events:

$$P(M_H|x_1 \dots x_n) = \prod P(M_H|x_i)$$

- Higgs mass is correctly reconstructed - modes of the probability distributions at correct Higgs masses (~2000 events, quite a high statistics)

Y-axis - logarithmic scale

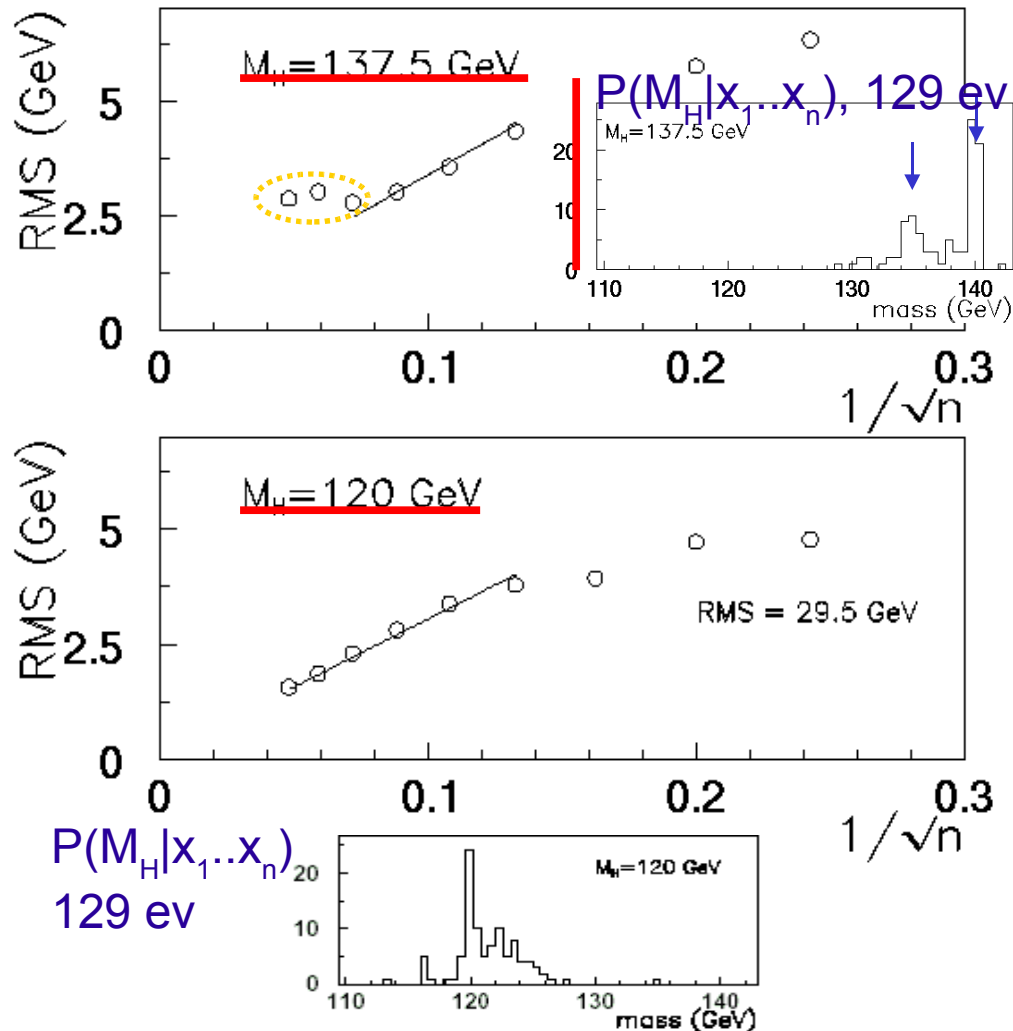
# Mass reconstruction for lower statistics



Reconstructed mass as a function of a number of events in the sample.

- Mean of the  $P(M_H | x_1 \dots x_n)$  distribution used as a mass estimate.
- Mass properly reconstructed (events with  $M_H = 137.5 \text{ GeV}$  were not used for training).
- "Edge effect" - a bias due to the weighted average taken over a finite Higgs mass range.

# Estimation of resolution



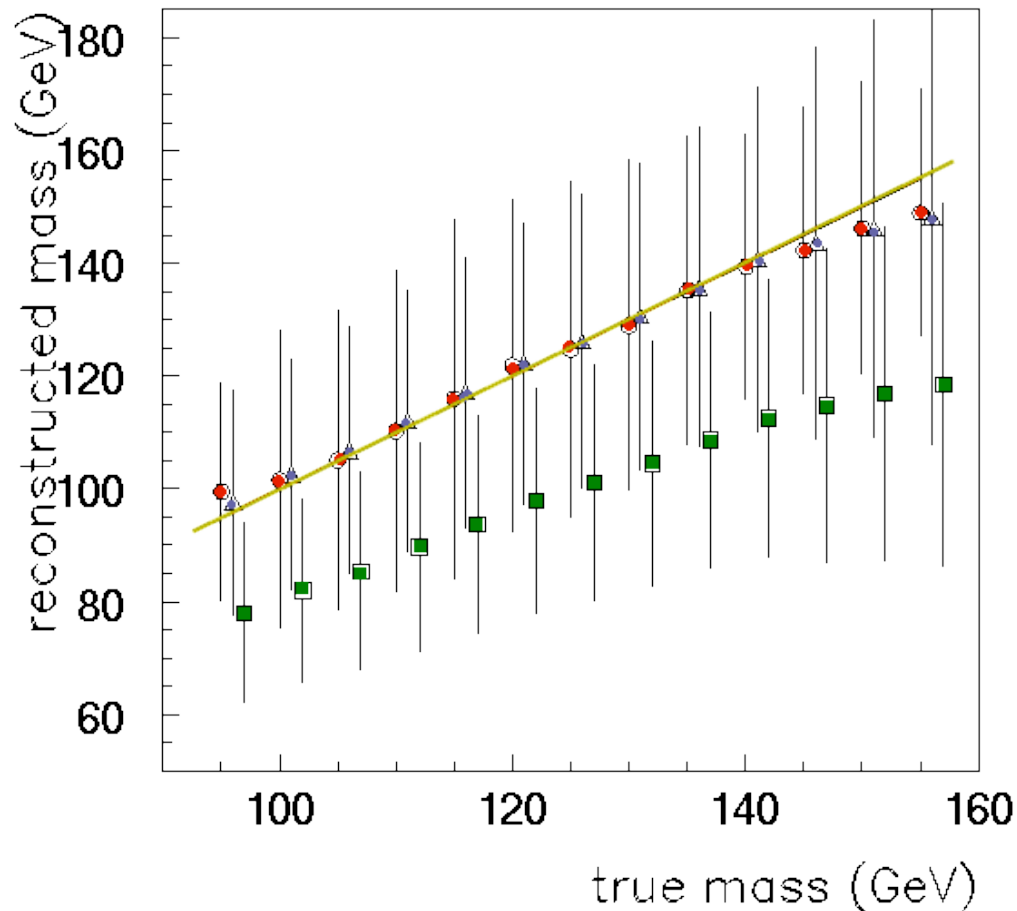
- Finding RMS of the estimated mass distribution for various sizes  $N$  of event samples.
- Data fit to the dependence:

$$RMS(n) = \frac{RMS(1)}{\sqrt{n}}$$

where  $RMS(1)$  is a single event RMS.

- Two peaks for  $M_H = 137.5$  GeV - the algorithm tends to return one of the neighbor masses. Effect could be minimized by generating more MC samples.

# Comparison with invariant mass method



- Red - Bayesian method.
- Green - mean of the inv. mass distribution.

$$M_{inv}^2 = 2 E_1 E_2 (1 - \cos \theta_{12})$$

- Blue - scaled mean of the invariant mass distribution.
- Error bars represent RMS
- Bayesian approach gives results very similar to the use of invariant mass, but no information encoded in the invariant mass formula are used.

# Conclusions

---

- Bayesian method works well and can be applied without knowing the functional dependence of particle properties on the measured quantities.
- No more information can be extracted from the  $E_1, E_2$  and  $\cos(\theta)$  beyond that encoded in the invariant mass.
- By adding additional variables it may be possible to improve the power of the Bayesian method.
- Neural network is a good fitting tool - unbinned multidimensional fit.

# Appendix

---



# Monte Carlo simulation

---

- All events were generated using CompHEP version 33.23, hadronized with Pythia version 6.158 and event book-keeping was done using STDHEP version 4.09.
- The detector simulation used PGS version 3.0 (Pretty Good Simulation, fast simulation package developed by John Conway et al.) with the following parameters:

- number of eta cells in calorimeter: 84
- number of phi cells in calorimeter: 64
- width of each cell in eta: 0.10
- width of each cell in phi: 0.10
- em calorimeter resolution: 0.20
- hadron calorimeter resolution: 0.80
- trigger MET resolution: 0.20
- calorimeter cell edge crack fraction: 0.05

- cluster seed tower threshold: 3.00
- shoulder tower threshold: 0.50
- cluster finder cone size: 0.70
- outer radius of tracking (m): 1.00
- magnetic field (T): 1.40
- sagitta resolution (m): 0.00004
- track efficiency: 0.98
- minimum track pt: 0.30
- tracking eta coverage: 2.00



# Rev. Thomas Bayes



**Thomas Bayes** (b. 1702, London - d. 1761, Tunbridge Wells, Kent), mathematician who first used probability inductively and established a mathematical basis for probability inference (a means of calculating, from the number of times an event has not occurred, the probability that it will occur in future trials). He set down his findings on probability in "Essay Towards Solving a Problem in the Doctrine of Chances" (1763), published posthumously in the *Philosophical Transactions of the Royal Society of London*.

Article on WEB: <http://www.uv.es/valencia7> (J. Bernardo)

Book: D.S. Sivia "Data Analysis - a Bayesian Tutorial"



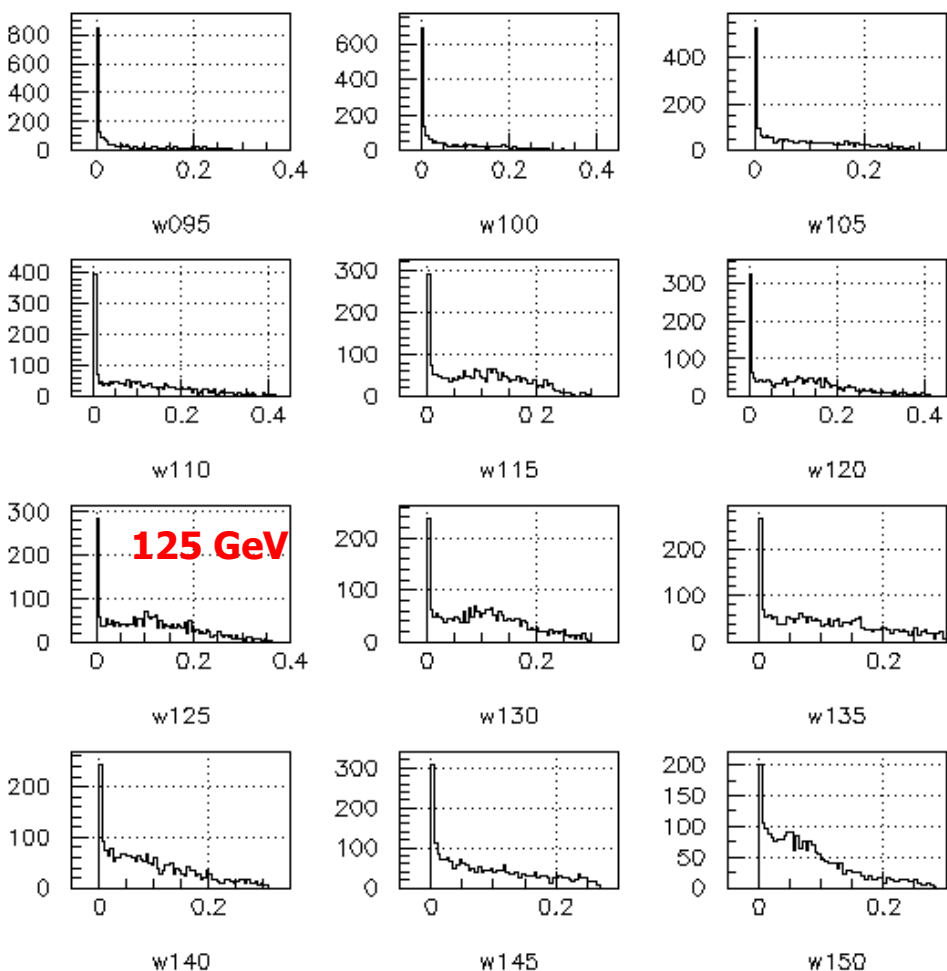
# Statistics of Bayes & Laplace

---

- Probability is a *degree-of-belief* or plausibility (not defined by frequency of occurrence).
- To measure the parameter  $Y$  having data  $X$  we look for the conditional probability  $P(Y/X)$ .
- In the analysis we do not need to know the functional dependence  $Y=f(X)$ , it is enough to know  $P(X/Y)$  and than obtain  $P(Y/X)$  from the Bayes theorem:  
$$P(\text{hypothesis}/\text{data}) \propto P(\text{data}/\text{hypothesis}) \times P(\text{hypothesis})$$



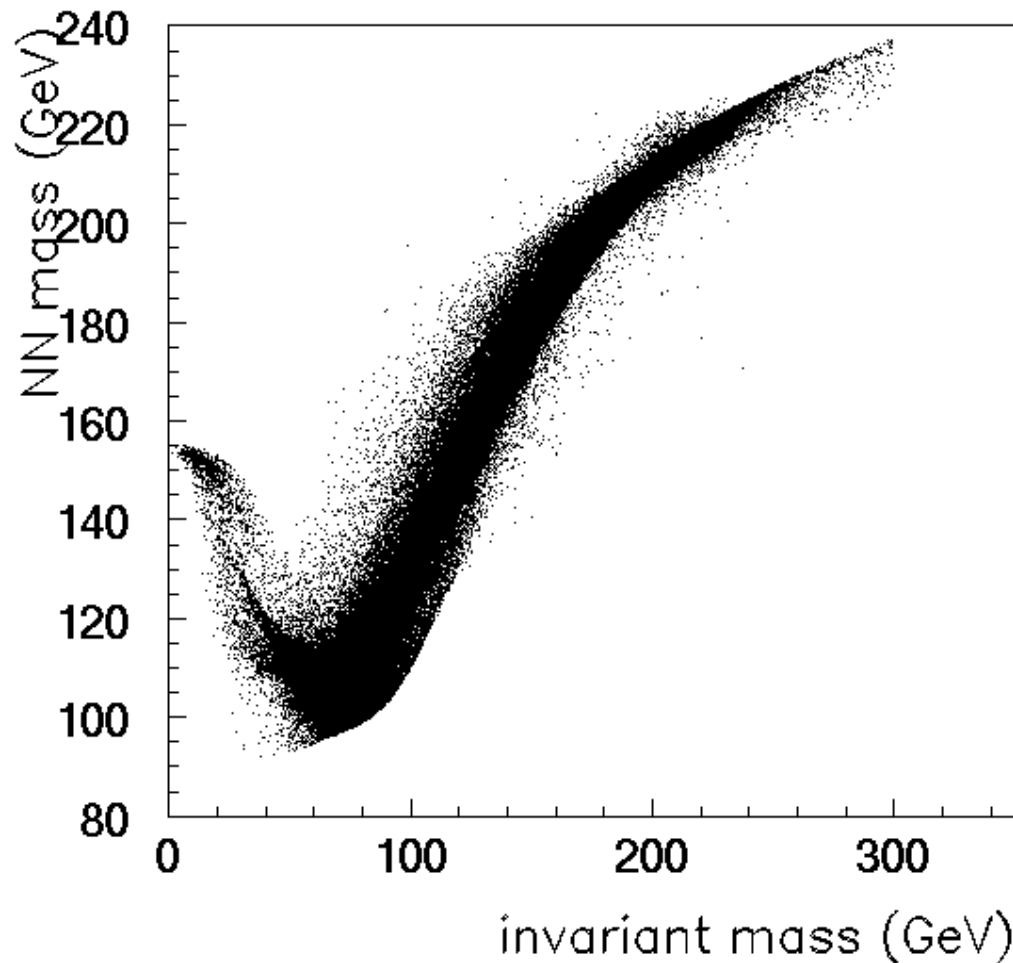
# Single event probabilities



Distributions of single events probabilities for  $M_H = 125$  GeV sample.

Highest probabilities for 125 GeV, as expected.

# Network returning Higgs mass



Higgs mass reconstructed by a neural network vs. the standard invariant mass - the "mass shifting" effect due to the  $\chi^2$  minimization is clearly visible.