Measurement of physical quantities in the Bayesian framework using neural networks

Marcin Wolter
Tufts University, Boston, USA
Institute of Nuclear Physics, Krakow, Poland

18 March 2002

Abstract

I study the Bayesian approach to the reconstruction of particle properties using as an illustration the measurement of the Higgs boson mass. Monte Carlo events generated with various Higgs masses are used to determine the probability distribution for each Higgs boson mass. The probability is a function of measured quantities, in this example energies of b-tagged jets as well as the the angle between them. The probability distributions are fitted using neural networks and the final Higgs mass probability distribution is obtained as a product of the single event distributions. The miscalibration of the measured quantities is automatically corrected by the probability distributions.
I also study a complementary method based on a neural network trained to return directly the Higgs mass.

Introduction

The precision of particle property reconstruction can be improved by using data analysis methods that better explore all the accessible information. The Bayesian approach, based on an interpretation of probability as a conditional measure of uncertainty, provides such an opportunity. In this method for every event the probability $P(m|x)$ of x belonging to class m is computed. In our example "class" m is the Higgs mass and x is a vector of variables. The event x is assigned to the class with the highest probability. For a sample of many events originating from the same, but unknown class, the probability is a product of the single event probabilities:

$$P(m|x_1, ..., x_n) \propto \left(\prod_{i=1}^{n} P(x_i|m) \right) \times P(m) \ ,$$

where $P(m)$ is a prior on the Higgs mass which we shall take to be constant. In the limit of a continuous probability function of the particle mass this method gives the best possible estimate of the mass, provided that the probability functions are well measured and the vector x describing the event contains all the necessary information. One should note, that no explicit knowledge of the functional dependence of the mass estimate (invariant mass) is needed.

We consider the reaction $H \to b\bar{b} \to 2\ jets$, in which the Higgs boson decays into two b-mesons producing two jets. The obvious and simplest estimate of the the Higgs boson mass is the invariant mass of the two jets. To calculate the mass estimate the energies of two jets and the angle between them are needed. The same three variables are used in the Bayesian analysis.

The energies of reconstructed jets are not calibrated consequently the invariant mass of two jets is shifted towards lower values (Fig. 1). In this initial study, neither physical nor combinatorial background is present. Only perfectly tagged b-jets are used.

Method

Monte Carlo events were generated using PGS simulation [1]. Thirteen samples with Higgs masses between 95 GeV and 155 GeV with 5 GeV increments were generated. For each of the above Higgs masses the probability function $P(x|m)$ (where x is a vector of two jet energies and cosine of the angle between them - e_1,e_2 and $\cos\theta$) is fitted using a neural network.

All the simulated samples are divided into two subsamples: one is used for training the neural network, the second for further analysis. Following Ref. [3] each of the samples was trained against a sample with a flat distribution in all three variables: e_1,e_2 and $\cos\theta$. The feed-forward network is trained to return "1" for the PGS Monte Carlo sample and "0" for the "flat" sample. The probability $P(x|m)$ is obtained from the output NN_{out} of the adequately trained neural network:

$$P(x|m) \propto \frac{NN_{out}}{1 - NN_{out}} \; .$$

The network consists of three input nodes, 50 nodes in the hidden layer and one output node. There are 13 independent networks, each of them trained using a sample with different Higgs mass. All fitted functions are later normalized, so the probabilities for different Higgs masses can be compared. Figure 1 shows the standard invariant mass compared with the invariant mass obtained from the fitted 3-dimensional function. The fit smoothes the distribution, as expected.

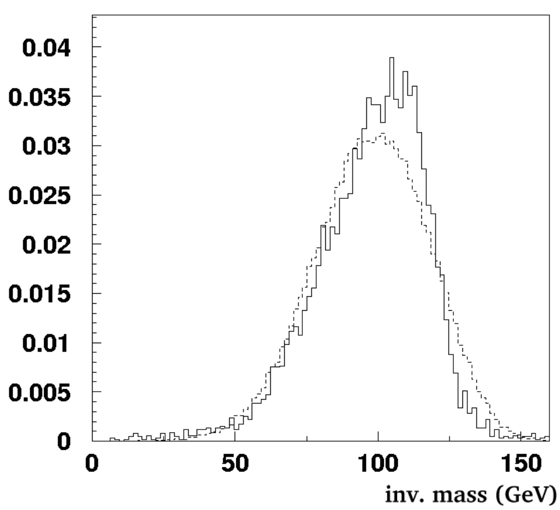

Figure 1: Standard invariant mass (solid line) compared with the invariant mass obtained from the Neural Network fit (for $M_H = 120\ GeV$ sample).

One advantage of a neural network fit is the fact that data are not binned, which dramatically improves the quality of the fit while fitting small samples. Also no analytical formula of the fitted function is needed. The complexity of the function shape is determined by the number of links (free parameters) in the neural network. The trained neural network can be easily converted into a C-language function, which later returns the probability at very low CPU cost.

For each event the trained neural networks return the set of probabilities corresponding to each Higgs mass. To obtain the mass estimate based on a sample of few events the probabilities should be multiplied. The mode of the distribution can be taken as an estimate of the Higgs mass. Instead of the mode the mean is an estimate. This estimate reduces fluctuations and improves the mass resolution.

Results

The probability distributions as functions of the Higgs mass are shown in Fig. 2 for samples of 2300 events generated with four different Higgs masses. On the Y-axis the $\ln(P(x|m))$ is plotted, therefore heights of the bins differ by orders of magnitude. The plot shows, that the maxima are located at the true Higgs mass values.

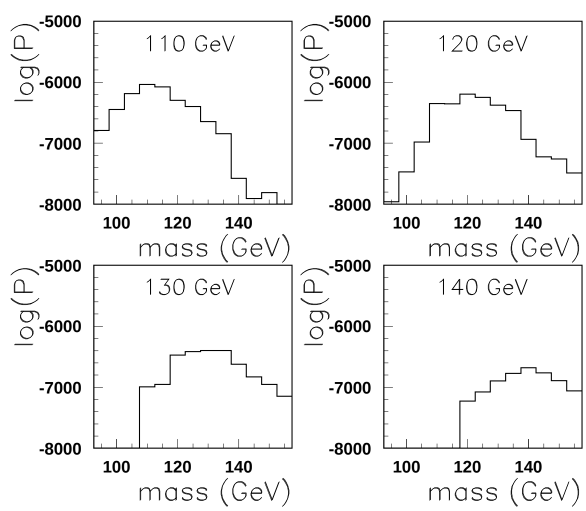

Figure 2: Probability distributions for the samples of 2000 events generated at four Higgs masses: 110 GeV, 120 GeV, 130 GeV and 140 GeV.

It is useful and necessary to test this method on a sample generated with a Higgs mass different from the masses used for training. Fig. 3 shows the reconstructed mass as a function of the number of events used for reconstruction (MC sample with Higgs mass 137.5 GeV). On the Y-axis the mean mass over many small subsamples is plotted. The bias at low sample size is an "edge effect" due to the weighted average being take over a finite Higgs mass range. The effect

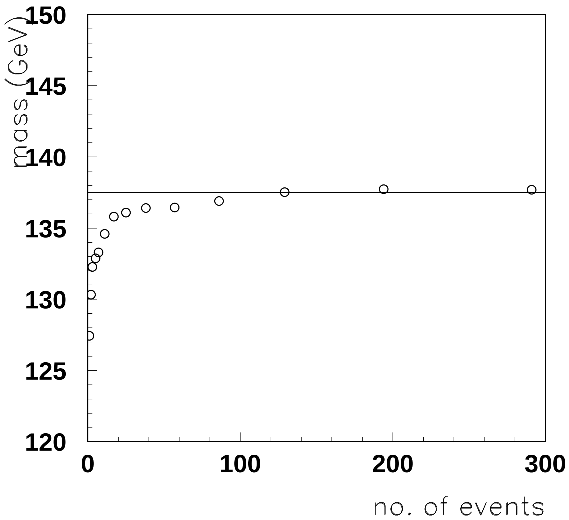

Figure 3: Mean reconstructed mass as a function of the sample size for a sample generated with Higgs mass of 137.5 GeV.

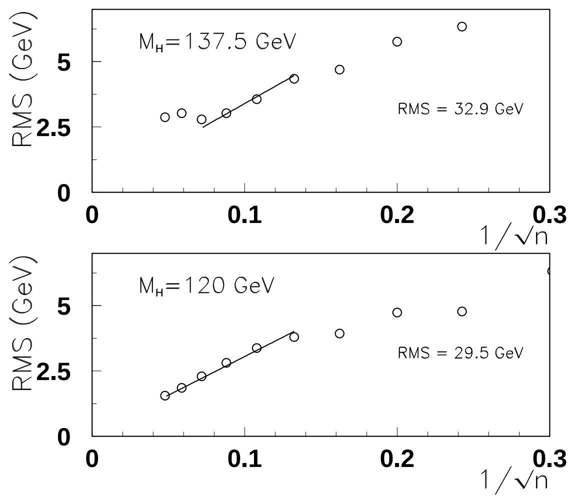

Figure 4: RMS as a function of $1/\sqrt{n}$, where n is a number of events in a sample. Upper plot shows the dependence for a sample with $M_H = 137.5\ GeV$, the lower for $M_H = 120\ GeV$.

can be reduced by using a wider Higgs mass range.

The resolution of the mass estimation is obtained by analyzing the root mean square (RMS) of the reconstructed mass distribution as a function of the number of events in the sample. It should scale according to the formula

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

where n is a number of events in a sample. Fig. 4 shows this dependence for samples generated with Higgs masses 120 GeV and 137.5 GeV. The violation of the $1/\sqrt{n}$ dependence at small sample size is due to the fact, that the RMS is limited by the Higgs mass range (i.e. 95 GeV to 155 GeV). For greater numbers of events in a sample some violation of this dependence for $M_H = 137.5\ GeV$ sample is observed. Since the closest generated Higgs masses are 135 GeV and 140 GeV, the reconstructed mass tends to be equal to one of them, and therefore the RMS is approximately half of the difference between them, i.e. 2.5 GeV (see Fig. 5). This effect can be reduced by generating MC samples with intermediate Higgs masses.

Fig. 6 shows the dependence of the reconstructed mass (weighted mean for a sample of 57 events) and single event RMS as a function of the true Higgs mass. The dependence is fairly linear and the mass is properly reconstructed. The results are compared to the mean and RMS of the invariant mass distribution, which gives the masses significantly lower than the

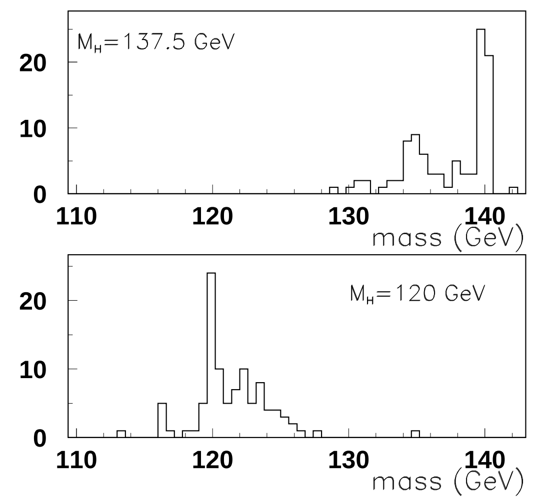

Figure 5: Reconstructed mass distribution for a sample size of 129 events and for $M_H = 137.5\ GeV$ (upper plot) and for $M_H = 120\ GeV$ (lower plot).

true Higgs boson mass. The invariant mass is scaled by a factor of 1.25 and compared to the results of the Bayesian method. The reconstructed masses and their errors are very similar for both methods, i.e. using Bayesian probabilities and the standard invariant mass approach with scaling.

These results show, that the method described here could be effective when the functional dependence of the reconstructed mass or other particle property is unknown.

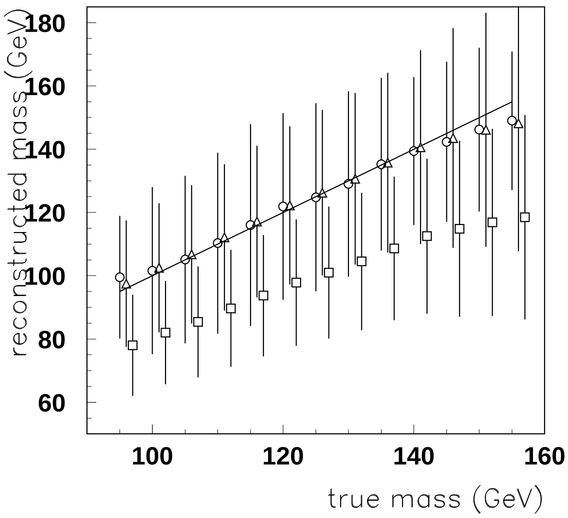

Figure 6: Reconstructed Higgs mass vs. the true Higgs mass for the Bayesian method (circles), mean of the invariant mass distribution (squares) and corrected mean of the invariant mass distribution (crosses). The error bars represent RMS of the distributions. The data points are shifted for better visualization.

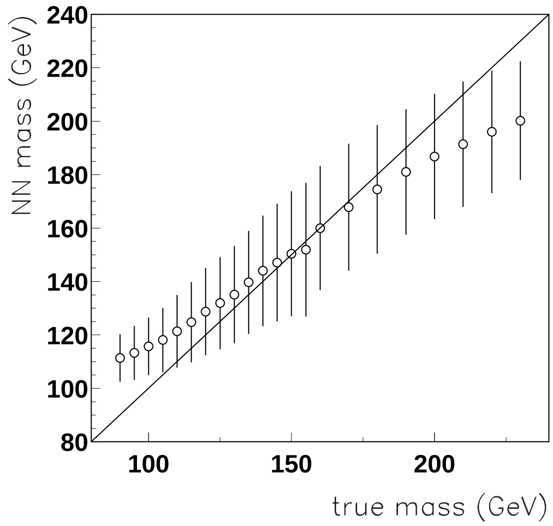

Figure 7: Reconstructed Higgs mass vs. the true Higgs mass for the neural network directly returning the Higgs mass. The error bars represent RMS of the distributions.

Neural network trained to return Higgs mass directly

A more straight forward approach to the particle mass reconstruction is to train a neural network to return directly the Higgs mass. This method requires training only one network and therefore requires much less CPU. A neural network with two hidden layers (10 nodes in each layer) was trained using the extended data sample (Higgs mass from 95 GeV up to 230 GeV in 10 GeV intervals) to reduce the edge effects.

The results are shown in Fig. 7. Also the samples not used in the neural network training are included (samples with Higgs mass 95 GeV, 105 GeV, 115 GeV, 125 GeV,135 GeV, 145 GeV and 155 GeV). The RMS errors are moderate, however the reconstructed masses differ from the expected values. They are shifted towards the middle of the spectrum, i.e. low masses are higher than expected and high masses are lower.

The training of the neural network minimizes the χ^2 difference between the expected and returned Higgs masses. However, due to the broad invariant mass distribution, and therefore the similarity of the events with different masses, the minimal χ^2 is obtained by narrowing the spectrum of reconstructed masses. It is clearly visible in Fig. 8, where the mass reconstructed by the neural network is plotted against the regular invariant mass. The events with very low invariant masses are pushed towards higher masses and events with high invariant mass are pushed towards lower masses. This, again, is an edge effect. A training sample with a much broader Higgs mass spectrum would be necessary to gain the linear dependence.

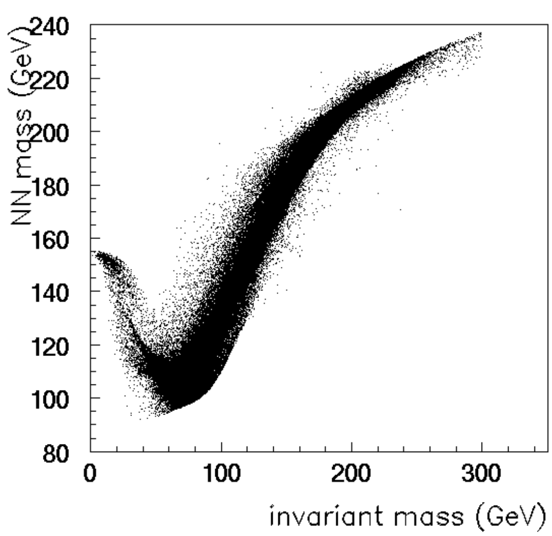

Figure 8: Higgs mass reconstructed by a neural network vs. the regular invariant mass.

Conclusions

The Bayesian approach to the problem of the reconstruction of the particle mass or other particle properties can be performed without any knowledge of the functional dependence of the particle property on measured quantities. However, for this method, as for other methods, a good Monte Carlo simulation of the physical processes and the detector is essential. In the example presented here, where the Higgs boson mass is measured, the method gives a mass resolution similar to the one obtained using standard invariant mass analysis. I conclude that no more information can be extracted from the jet energies and angle between them beyond that encoded in the invariant mass. By adding additional variables it may be possible to improve the power of the Bayesian method.

It was also shown, that the neural network is an excellent tool not only for signal and background discrimination, but also to perform multidimensional unbinned fits. The neural network, when already trained, returns the results and nearly no CPU cost and can be simply incorporated as a function into the analysis code.

Acknowledgments

Author would like to thank Pushpalatha Bhat, Harrison Prosper and other members of the Fermilab "Run II Advanced Analysis Group" for their interest in the study presented, important discussions and comments.

References

[1] PGS (Pretty Good Simulator formerly known as SHW) is a fast approximate detector simulation program, developed by John Conway and colleagues
http://www.physics.rutgers.edu/~jconway

[2] SNNS - Stuttgart Neural Network Simulator, v.4.2.
http://www-ra.informatik.uni-tuebingen.de/SNNS/

[3] Llus Garrido and Aurelio Juste *On the determination of probability density functions by using Neural Networks* , Computer Physics Communications 115 (1998) p. 25