Extraction of Drell-Yan Angular Coefficients using Neural Network-based Classifiers

Dinupa Nawarathne

New Mexico State University
Representing the E-906/SeaQuest Collaboration

June 20, 2023

Table of Contents

Likelihood Ratio Test

- The likelihood ratio test is a highly effective method for assessing the goodness of fit.
- Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. To decide between two simple hypotheses $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, we define the likelihood ratio:

$$\lambda(x_1, x_2, \ldots, x_n) = \frac{L(x_1, x_2, \ldots, x_n; \theta_0)}{L(x_1, x_2, \ldots, x_n; \theta_1)}$$

- To perform a likelihood ratio test, we choose a constant c. We reject H_0 if $\lambda < c$ and accept it if $\lambda \geq c$. The value of c can be chosen based on the desired significance level α.

Likelihood Ratio Test

- Neural networks excel at approximating functions, making them ideal for use as higher-dimensional likelihood functions.
- Our goal is to train the neural network to classify samples accurately. Specifically, we aim to classify samples $\omega_{0i} \in \Omega_0$ as $y = 0$ and $\omega_{1i} \in \Omega_1$ as $y = 1$, regardless of the parameter θ.
- Subsequently, we can utilize the trained neural network to estimate any unknown parameter $\theta_?$ by employing the gradient descent algorithm.[1]

[1] A. Andreassen, B. Nachman, *Phys. Rev. D* **101**, 091901, arXiv: 1907.08209 (hep-ph) (2020).

Gaussian Example

- Suppose we have a Gaussian distribution $\mathcal{N}(\mu_?, 1)$, where $\mu_?$ is the known parameter.
- Our training strategy is to train the neural network to classify samples drawn from two Gaussian distributions: $x_0 \sim \mathcal{N}_0(0, 1)$ with the label $y = 0$, and $x_1 \sim \mathcal{N}_1(\mu, 1)$ with the label $y = 1$, where μ can take any value within a specified range.
- We construct the neural network with three hidden layers, each layer consisting of 50 nodes. The layers are activated using the Rectified Linear Unit (ReLU), while the final layer is activated by the Sigmoid function.
- The neural network was trained for 200 epochs, employing early stopping with a patience of 10, to minimize the binary cross-entropy loss

Gaussian Example

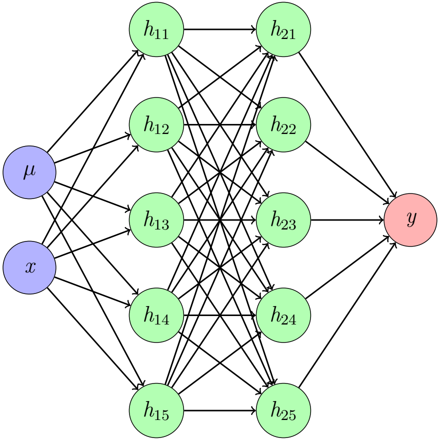

Figure 1: Example of a neural network architecture: The input layer is depicted with blue nodes, the hidden layers are represented by green nodes, and the output layer is indicated by a red node.

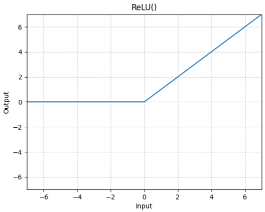

Figure 2: ReLU activation function.

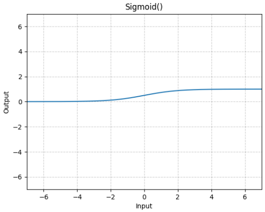

Figure 3: Sigmoid activation function.

Gaussian Example

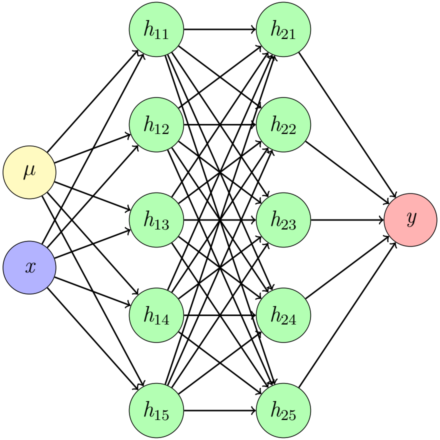

Figure 4: In the fitting step, the only trainable parameter is the yellow node representing μ.

- To perform the fitting algorithm, we fix the weights and biases of the trained neural network. Then, we optimize the μ parameter by using the gradient descent algorithm to find the optimal value μ_{opt}.

Gaussian Example

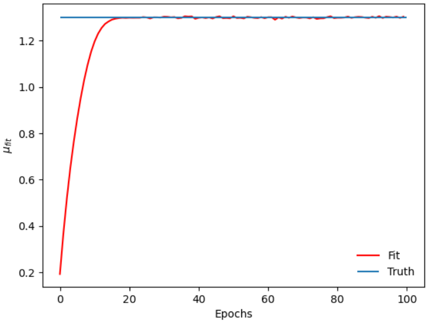

Figure 5: We tested the algorithm with $\mu_? = 1.3$. As depicted in the plot, the optimal value for μ converges to the correct value as the epochs progress.

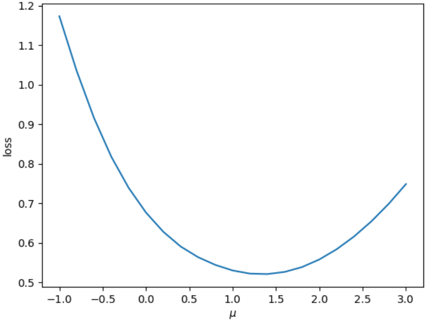

Figure 6: The loss reaches its minimum value at the optimal value of μ.

Drell-Yan Cross Section

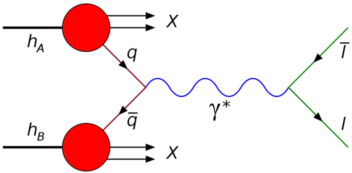

Figure 7: Drell-Yan process.

$$\frac{d\sigma}{d\Omega} \propto 1 + \lambda cos^2\theta + \mu sin2\theta cos\phi + \frac{\nu}{2} sin^2\theta cos2\phi$$

- In the ”naive” Drell-Yan model, where we ignore the transverse momentum of the quark and assume no gluon emission, we have $\lambda = 1$, $\mu = \nu = 0$.
- Measuring the $\cos 2\phi$ angular dependence can be used to extract the Boer-Mulders (BM) function.
- BM function describes the transverse-polarization asymmetry of quarks within an unpolarized hadron and results from the coupling between the transverse momentum and transverse spin of the quarks inside the hadron.

FermiLab E906/SeaQuest Experiment

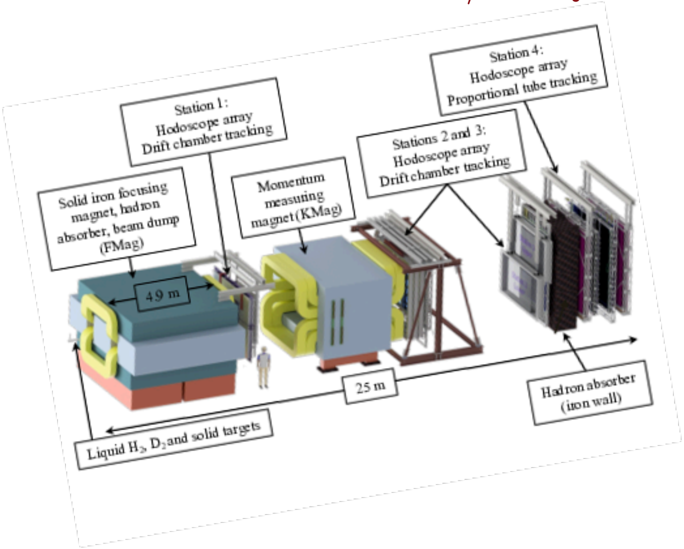

Figure 8: FermiLab E906/SeaQuest spectrometer.

Applying the Fitting Algorithm to E906 MC Data

- We generated the Monte Carlo (MC) data using the PYTHIA generator. The generated events were then passed through the E906 detector simulation to obtain the reconstructed detector information. As input features for the neural network, we use the mass, p_T, x_F, ϕ, and $\cos\theta$.
- We sample the values of λ, μ and ν uniformly from the ranges of $(0.5, 1.5)$, $(-0.5, 0.5)$, and $(-0.5, 0.5)$, respectively.[2]
- The neural network consists of five hidden linear layers, each containing 64 nodes. The ReLU function is used to activate the hidden layers, along with batch normalization layers. The final output is passed through a Sigmoid activation function.

[2] L. Y. Zhu *et al.*, *Phys. Rev. Lett.* **99**, 082301, arXiv: hep-ex/0609005 (2007).

Applying the Fitting Algorithm to E906 MC Data

- The neural network was trained for 200 epochs, employing early stopping with a patience of 20, to minimize the binary cross-entropy loss.
- We conducted a sanity check using five different combinations of λ, μ, and ν. The fitted values are presented in table 1.
- In the five test samples, we were able to extract the injected parameters within a ± 1.5 standard deviation (σ) interval.

Applying the Fitting Algorithm to E906 MC Data

Combination	Coefficient	Injected	Fitted
1	λ	0.84	0.876 ± 0.208
	μ	0.26	0.234 ± 0.054
	ν	-0.34	-0.299 ± 0.052
2	λ	1.33	1.134 ± 0.151
	μ	0.17	0.146 ± 0.050
	ν	-0.34	-0.281 ± 0.043
3	λ	1.12	1.242 ± 0.181
	μ	-0.27	-0.211 ± 0.088
	ν	-0.24	-0.236 ± 0.071
4	λ	0.62	0.888 ± 0.282
	μ	-0.32	-0.232 ± 0.091
	ν	0.18	0.147 ± 0.055

Table 1: Table showing the fitted values of λ, μ, and ν using the gradient descent algorithm.

Summary

- Neural networks can be used as multi-dimensional likelihood functions, and we can utilize likelihood ratio test to extract the optimal parameters for the Drell-Yan angular distribution.
- Measuring the $\cos 2\theta$ angular dependence with higher accuracy is important for extracting the Boer-Mulders (BM) functions.
- BM function describes the transverse-polarization asymmetry of quarks within an unpolarized hadron and results from the coupling between the transverse momentum and transverse spin of the quarks inside the hadron. This function serves as a useful tool for unraveling the structure of hadrons.
- Our plan is to use this high-dimensional fitting algorithm to extract the Drell-Yan angular coefficients from the E906/SeaQuest data with higher accuracy.