Probing BSM effects in $e^+e^- \to W^+W^-$ with Machine Learning

Shengdu Chai

Fudan University, Physics Department

February 13, 2023

Current work with Jiayin Gu and Lingfeng Li

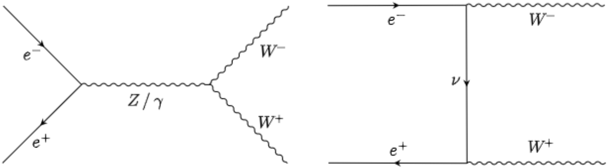
Motivation

- In many cases, the new physics contributions are sensitive to the differential distributions.
  - How to extract information from the differential distribution?
  - If we have the full knowledge of $\frac{d\sigma}{d\Omega}$, matrix element method, Optimal Observables, etc. can be used.
- The ideal $\frac{d\sigma}{d\Omega}$ we can calculate is not the $\frac{d\sigma}{d\Omega}$ we can measure.
  - Detector acceptance, measurement uncertainties, ISR/beamstrahlung ...
  - In practice we only have MC samples, not the analytical form, $\frac{d\sigma}{d\Omega}$.
- How machine learning works?
  - Black box.
  - Input: MC samples, output: likelihood ratio.

Why $e^+e^- \to W^+W^-$

- Focusing on $e^+e^- \to W^+W^-$.
- An important part of the precision measurement program.
- Connected to the higgs couplings in the SMEFT.
- Can be measured very well at Higgs factories.

EFT Parameterization(TGCs)

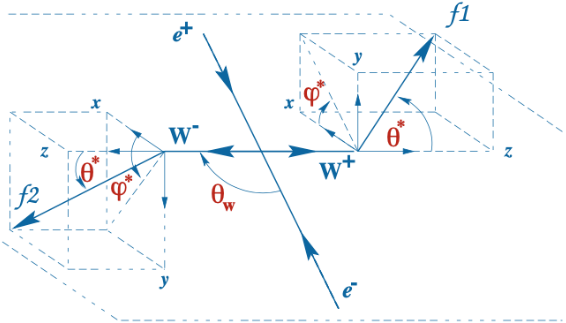


- Focusing on tree-level CP-even dimension-6 contributions.
- $e^+e^- \to W^+W^-$ can be parameterized by

$$\delta g_{1z}, \delta\kappa_\gamma, \lambda_\gamma, \delta g_L^{Ze}, g_R^{Ze}, \delta g_L^{Wl}, \delta_m$$

- m_W is better constrained, so we can simply set $\delta_m = 0$.

$e^+e^- \to W^+W^-$ with Histogram



- New physics contribution are sensitive to the differential distributions.
    - One could do a fit to the binned distributions of all angles.
    - Not the most efficient way of extracting information.
    - Correlations among angles are sometimes ignored.

$e^+e^- \to W^+W^-$ with Optimal Observable

- What are Optimal Observables?

  Diehl, M., Nachtmann, O., 1994. Zeitschrift Für Physik C Part Fields 62, 397–411.

  - In the limit of large statistics (everything is Gaussian) and small parameters (linear contribution dominates), the best possible reaches can be derived analytically!

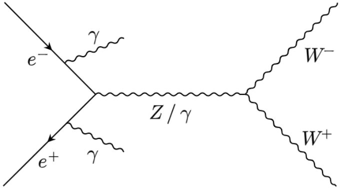
$$\frac{d\sigma}{d\Omega} = S_0 + \Sigma_i S_{1,i} g_i, \qquad c_{ij}^{-1} = \int d\Omega \frac{S_{1,i} S_{1,j}}{S_0} \cdot \mathcal{L}$$

- The optimal observable is a function of 5 angles and is given by $\mathcal{O}_i = \frac{S_{1,i}}{S_0}$.

Systematic Effects

- Initial state radiation

[2108.10261] Frixione, Mattelaer, Zaro, Zhao



$$\Gamma_{e^\pm/e^\pm}(z) = \frac{e^{3\beta/4-\gamma_E\beta}}{\Gamma(1+\beta)}\beta(1-z)^{\beta-1} - \frac{\beta}{2}h_1(z) - \frac{\beta^2}{8}h_2(z),$$

$$h_1(z) = 1+z,$$

$$h_2(z) = \frac{1+3z^2}{1-z}\ln(z) + 4(1+z)\ln(1-z) + 5 + z,$$

Systematic Effects

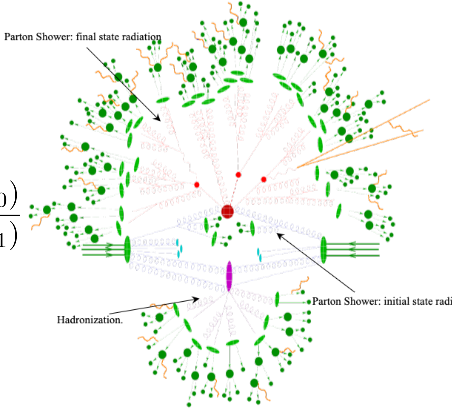
- Jet smearing
  - Photon and neutral hadron energy resolutions.
  - The system error are assumed to be Gaussian distributions.
- Detect effects
  - Final state jets can not be distinguished.
  - Neutrino cannot be directly measured.

Likelihood Inference

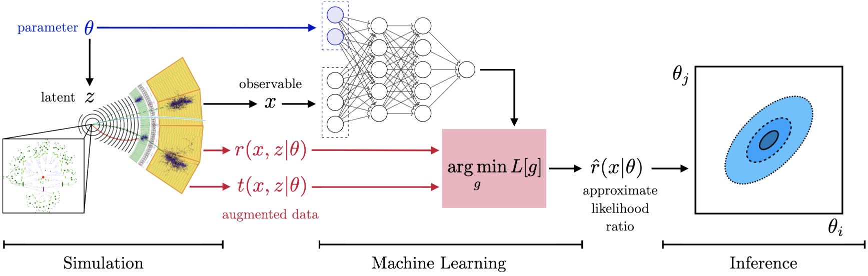
- Neyman-Pearson lemma says the best statistics to test new physics is the likelihood ratio given data x and theory parameters θ_1 and θ_0

$$\hat{r}(x|\theta_0,\theta_1) = \frac{p(x|\theta_0)}{p(x|\theta_1)} = \frac{\int dz\, p(x,z|\theta_0)}{\int dz\, p(x,z|\theta_1)}$$

- The key thing is $\hat{r}(x|\theta_0,\theta_1)$.
- Analytical methods always computational consuming and ignore systematic effects.



Likelihood Inference



- Johann Brehmer et al. develop new simulation-based inference techniques that are tailored to the structure of particle physics processes. [1805.00013]Brehmer, Cranmer, Louppe, Pavez.
- Machine Learning method can extract more information from x to predict the likelihood ratio.

Particle-Physics Structure

- The likelihood function can be written as

$$p(x|\theta) = \int dz \; p(x, z|\theta) = \int dz \; p(x|z)p(z|\theta)$$

- Here $p(z|\theta) = 1/\sigma(\theta) d\sigma/dz$ is the parton level density distribution.
- $p(x|z)$ describes the probabilistic evolution from the parton-level four-momenta to observable particle properties

$$p(x|z) = \int dz_d \int dz_s \int dz \; p(x|z_d)p(z_d|z_s)p(z_s|z)$$

Particle-Physics Structure

- We can extract more information from the simulator by defining joint likelihood ratio and joint score

$$r(x, z|\theta_0, \theta_1) = \frac{p(x|z)p(z|\theta_0)}{p(x|z)p(z|\theta_1)} = \frac{p(z|\theta_0)}{p(z|\theta_1)}$$

$$\alpha(x, z|\theta_0, \theta_1) = \nabla_{\theta_0} r(x, z|\theta_0, \theta_1)|_{\theta_0=\theta_1}$$

- The loss function is

$$\mathcal{L}[\hat{g}(x)] = \int dxdz\; p(x, z|\theta)|g(x, z) - \hat{g}(x)|^2$$

- The loss function is minimized when $\hat{g}(x) = \frac{1}{p(x|\theta)} \int dz\, p(x, z|\theta) g(x, z)$
- $g(x, z) = r(x, z|\theta_0, \theta_1)$, and $\theta = \theta_1$,$\hat{g}(x) = \hat{r}(x|\theta_0, \theta_1)$.

ML Algorithm: ALICE

- Approximate likelihood with improved crossentropy estimator
- Directly predict the likelihood ratio.
- Loss function $\mathcal{L}$ is

$$\mathcal{L}(\hat{s}) \propto \sum_x [s(x,z|\theta_0,\theta_1)\log(\hat{s}(x)) + (1 - s(x,z|\theta_0,\theta_1))\log(1-\hat{s}(x))]$$

- Here $s(x,z|\theta_0,\theta_1) = \frac{1}{1+r(x,z|\theta_0,\theta_1)}$.
- $\hat{r}(x|\theta_0,\theta_1)$ can be reconstructed by $\hat{s}(x) = \frac{1}{1+\hat{r}(x|\theta_0,\theta_1)}$.

ML Algorithm: SALLY

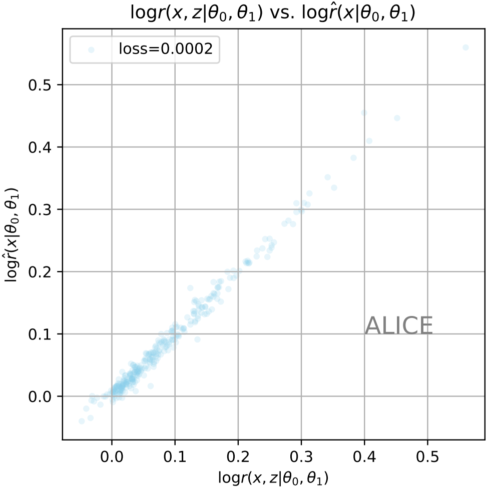
- Score approximates likelihood locally
- Likelihood ratio can also be parameterized by Wilson coefficients.

$$\hat{r}(x|\theta) = 1 + \sum_i \hat{\alpha}_i(x)\theta_i$$

- And we can predict α_i term as well.
- Loss function $\mathcal{L}$ is

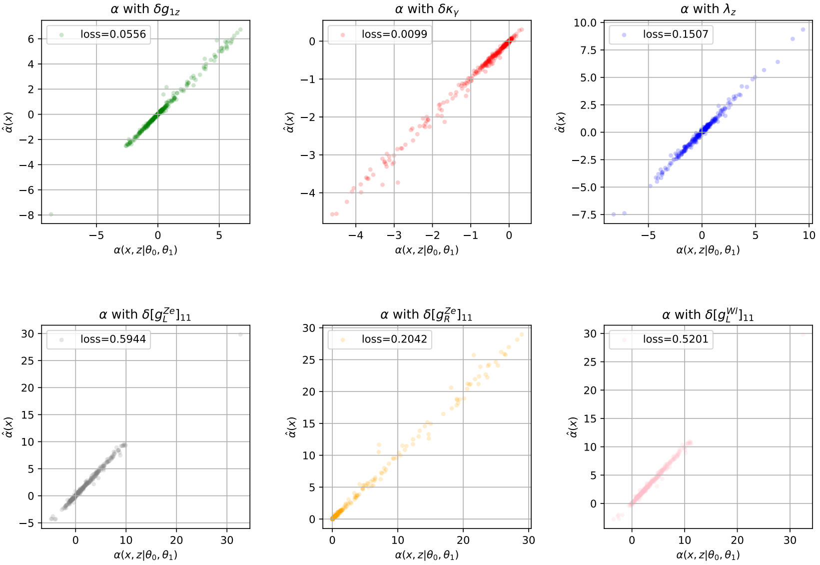
$$\mathcal{L} \propto \sum_i |\hat{\alpha}_i(x) - \alpha_i(x, z|\theta_0, \theta_1)|^2$$

Prediction of Likelihood Ratio:ALICE

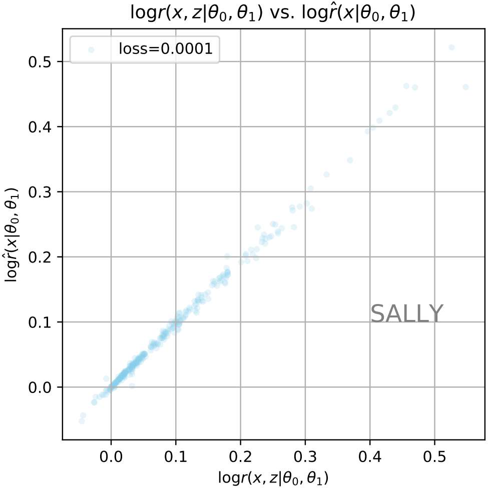


- ALICE method offers a precise way to predict the likelihood ratio directly.

Prediction of Likelihood Ratio:SALLY

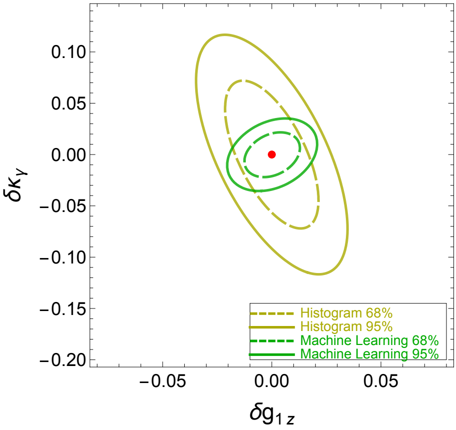


Prediction of Likelihood Ratio:SALLY



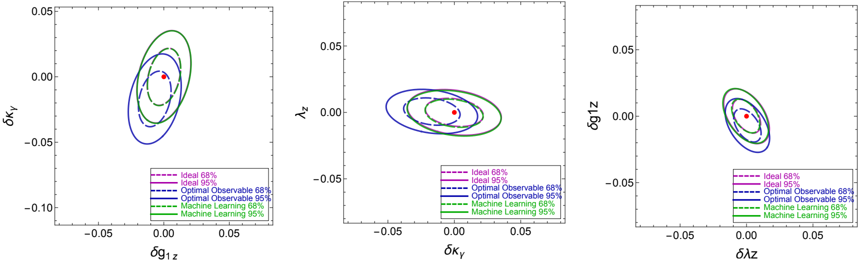
- We can construct the $\hat{r}(x|\theta_0, \theta_1)$ by predicting the alpha term and give an analytical expression of $\hat{r}(x|\theta_0, \theta_1)$.

Estimation of the Boundary:Compared with Histogram



- No bias.
- Precise bounds along individual directions.
- Weak constraints in other directions.

Estimation of the Boundary:Compared with OO



- Once you get the $\hat{r}(x|\theta_0, \theta_1)$, $\chi^2 = -2\sum_i \log(\hat{r}(x_i|\theta_0, \theta_1))$.
- Semileptonic channel, jet smearing + ISR, 3-aTGC fit
  - Naively applying optimal observables could lead to a large bias.
  - It is easier for machine learning to take care of systematic effects.

Conclusion

- Future colliders will generate large amount of data, ML will benefit it a lot.
- By machine learning, we can construct 6D likelihood ratio to improve the global fit result.
- Machine Learning can easily take care of systematic effects as long as the MC simulation is accurate.
- Machine learning is (likely to be) the future.

Thanks!

Backup Slides: $e^+e^- \to W^+W^-$ parameterization

$$\begin{aligned}\mathcal{L}_{\rm TGC} = ig\{&(W^+_{\mu\nu}W^{-\mu} - W^-_{\mu\nu}W^{+\mu})[(1+\delta g_{1z})c_\theta Z^\nu + s_\theta A^\nu] \\ &+ \frac{1}{2}W^+_{[\mu,}W^-_{\nu]}[(1+\delta\kappa_z)c_\theta Z^{\mu\nu} + (1+\delta\kappa_\gamma)s_\theta A^{\mu\nu}] \\ &+ \frac{1}{m_W^2}W^{+\nu}_\mu W^{-\rho}_\nu(\lambda_z c_\theta Z_\rho^{\ \mu} + \lambda_\gamma s_\theta A_\rho^{\ \mu})\}\end{aligned}$$

- Imposing Gauge invariance one obtains $\delta\kappa_Z = \delta g_{1,Z} - t^2_{\theta_w}\delta\kappa_\gamma$ and $\lambda_Z = \lambda_\gamma$
- $\delta g_{1z}, \delta\kappa_\gamma, \lambda_\gamma, \delta g_L^{Ze}, g_R^{Ze}, \delta g_L^{Wl}, \delta_m$