notes meeting thesis

tommaso perniola

November 2025

1 Additional Notes on GMMs and Distributional Modeling

1.1 Log-Likelihood

For probabilistic models, the log-likelihood quantifies how plausible the observed data are under the model. For a Poisson regressor, the log-likelihood of a single observation is:

$$\log p(y \mid \mu) = y \log \mu - \mu - \log(y!),$$

where y is the observed count and μ is the model prediction (the Poisson mean). Evaluating and plotting the log-likelihood is essential: classical metrics such as RMSE or MAE only assess mean error, while the log-likelihood reveals distributional mismatch (e.g. incorrect variance or shape). In the GRB background task, this is particularly relevant because we aim to model the entire count distribution, not only its expectation.

1.2 Interpolation Setting

In this project, the train–test split is random, so both sets are sampled from the same distribution. The learning task is therefore interpolation: the model is expected to reconstruct smooth mappings within the observed support rather than extrapolate to unseen orbital configurations or background regimes. This assumption is compatible with KDE, GMMs and Poisson regression.

1.3 Why GMMs

Gaussian Mixture Models (GMM) We consider GMMs because a single continuous distribution isn't always a good fit for our data. By using a mixture, we can fit each gaussian component separately and see which combination models the data better.

1.4   Enforcing Positivity

Photon count rates are strictly positive. However, Gaussian-based models (including GMMs or neural regressors with linear outputs) may produce negative predictions, which are physically meaningless. To enforce positivity, the raw network output z is passed through a transformation:

$$softplus(z) = \log(1 + e^z), \qquad or \qquad \exp(z).$$

The exponential parameterization is common when modeling scale parameters; in this case the model predicts $\tilde{\mu} = \log(\mu)$ and the final rate is obtained via $\mu = e^{\tilde{\mu}}$. These transformations guarantee positivity and stabilize optimization.

1.5   Residual Plots as Diagnostic Tools

Residuals are defined as $r_i = y_i - \hat{y}_i$ and are useful to identify bias, heteroscedasticity or systematic trends. However, they provide only qualitative hints. In distributional modeling (e.g. Poisson or Gaussian mixtures), small residuals do not guarantee that the assumed noise model is correct. A model may fit the mean well yet produce incorrect variance or density shape. Therefore, residual plots should be treated as a preliminary diagnostic rather than a final evaluation tool.

1.6   GMM Behaviour and Symmetry

Gaussian Mixture Models do not "overfit" in the same manner as flexible neural networks. Most issues arise from symmetry and component identifiability: mixture components may overlap, swap, or collapse depending on initialization. As a result, unexpected behaviour in GMMs typically reflects optimization difficulties or symmetry-related degeneracies rather than true overfitting.

1.7   SHAP and Output Transformations

When applying SHAP for interpretability, it is important to account for output transformations. SHAP explains the model in terms of its *final* output. If the model internally predicts a transformed quantity (e.g. a log-rate) and then exponentiates it, SHAP will attribute importance with respect to the exponential output. Since the exponential amplifies small changes in the latent space, feature attributions may become distorted. A common workaround is to compute SHAP values on the pre-transformation output (e.g. the predicted log-rate).