# Statistics and Data Analysis

## Evaluation Exercise for Unfolding Part

Simulate and solve an unfolding problem using the slides shown during the lectures (available at `Unfolding_pdf`) and the following steps:

1. Generate the input elements for the unfolding problem and compute non-regularized solutions:

   (a) Consider equal range and number of bins ($M = N$) for the true ($\boldsymbol{\mu}$) and measured distributions ($\boldsymbol{\nu}$), respectively. Construct the (square) migration matrix $R$ by assuming a (e.g.) Gaussian estimator with no bias and a resolution of 2 times the width of the bins of $\boldsymbol{\mu}$.

   (b) Construct the distribution $\boldsymbol{\mu}$ using a true p.d.f. ($f_{\text{true}}$) given by (e.g.) the sum of two Gaussian functions.

   (c) Construct the distribution $\boldsymbol{\nu}$ by multiplying the migration $R$ matrix to the vector $\boldsymbol{\mu}$.

   (d) Generate randomly the distribution $\boldsymbol{n}$ by assuming uncorrelated Poisson distributed measurements and no background.

   (e) Compute the exact solution $\boldsymbol{\mu}_0$ using the inverse of the migration matrix.

   (f) Compute (numerically) the maximum likelihood or minimum least-squares solution and compare it to the exact solution.

   (g) Compute a solution using the correction factors method.

   (h) Compute the solution for $f_{\text{true}}$ using the forward unfolding method (e.g. for the double-Gaussian case, the means, variances and normalizations are free parameters).

2. Using the elements obtained in Exercise 1, compute a regularized solution:

   (a) Consider the migration matrix $R$ and the distribution of measured values $\boldsymbol{n}$ produced in Exercise 1 (as in a real experiment, we assume we do not know $\boldsymbol{\mu}$ or $\boldsymbol{\nu}$, but we are going to estimate them).

   (b) Chose a regularization method (Tykhonov, maximum entropy,...).

   (c) Maximize Equation 7 of the slides systematically for a set of values of $\alpha$ (in practice we maximize Equation 5 keeping one $\boldsymbol{\mu}$ component dependent to the rest to keep condition $\nu_{\text{tot}}(\boldsymbol{\mu}) = n_{\text{tot}}$).

   (d) For each considered $\alpha$ value, compute the values of the four quantities used to chose the optimal $\alpha$ (MSE, MSE', $\chi^2_{\text{eff}}$ and $\chi^2_b$). Plot them as a function of $\Delta \log L$.

   (e) For each optimal $\alpha$ value, plot the estimated true distribution $\hat{\boldsymbol{\mu}}$ (with errors), together with the true distribution $\boldsymbol{\mu}$ and the measured distribution $\boldsymbol{n}$.

   (f) For each optimal $\alpha$ value, plot the estimated biases for the $\hat{\boldsymbol{\mu}}$ solution (with errors).