Calculation of Error Bars on the χ^2 Graph

To visualize the uncertainty in the cross-validation performance of different models, we compute error bars for the chi-squared values across multiple trials. These error bars represent the standard deviation of χ^2 values obtained from repeated experiments with different random seeds.

Let n_{trials} be the number of independent trials (datasets generated with different seeds). For each model complexity m (e.g., polynomial degree), we compute the chi-squared values:

$$\left(\chi_A^2\right)^{(i)}(m), \quad \left(\chi_B^2\right)^{(i)}(m), \quad \text{for } i = 1, \ldots, n_{\text{trials}}.$$

The average chi-squared values are:

$$\langle \chi_A^2(m) \rangle = \frac{1}{n_{\text{trials}}} \sum_{i=1}^{n_{\text{trials}}} \left(\chi_A^2\right)^{(i)}(m), \qquad \langle \chi_B^2(m) \rangle = \frac{1}{n_{\text{trials}}} \sum_{i=1}^{n_{\text{trials}}} \left(\chi_B^2\right)^{(i)}(m).$$

The corresponding standard deviations (used as vertical error bars) are:

$$\sigma_{\chi_A^2}(m) = \sqrt{\frac{1}{n_{\text{trials}} - 1} \sum_{i=1}^{n_{\text{trials}}} \left((\chi_A^2)^{(i)}(m) - \langle \chi_A^2(m) \rangle \right)^2},$$

$$\sigma_{\chi_B^2}(m) = \sqrt{\frac{1}{n_{\text{trials}} - 1} \sum_{i=1}^{n_{\text{trials}}} \left((\chi_B^2)^{(i)}(m) - \langle \chi_B^2(m) \rangle \right)^2}.$$

These values are plotted as vertical error bars around the mean χ^2 values for each degree m.