Principle of Maximum Likelihood Estimation (MLE)

In point-estimate ML, the Mean Squared Error (MSE) is commonly adopted as the evaluation metric, which quantifies the average squared deviation between deterministic predictions and observed values. However, MSE is limited to assessing single-value outputs and fails to account for uncertainty in probabilistic predictions. The MSE is mathematically defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{i,pred})^2 \tag{1}$$

where n denotes the total number of samples, y_i and $y_{i,pred}$ represent the observed (true) value and the predicted value of the i-th sample, respectively.

The NGBoost model estimates a parametric probability distribution $p(x|\theta)$ conditioned on input features $\{x_1, x_2, \dots, x_n\}$, necessitating evaluation metrics that quantify the discrepancy between the predicted distribution and observed data. NGBoost provides two probabilistic evaluation metrics: the negative log-likelihood (NLL) and the Continuous Ranked Probability Score (CRPS). The NLL is adopted in this study to assess model performance, as it holistically evaluates the calibration of the predicted distribution (shape, central tendency, and dispersion). The NLL is an optimization-friendly transformation of the Maximum Likelihood Estimation (MLE) principle, converting the likelihood maximization problem into a loss minimization task. The theoretical foundation of MLE is described as follows:

Given a parametric probability distribution $p(y_i|\theta, x_i)$, the likelihood function $\ell(\theta)$ can be defined as:

$$\ell(\theta) = \prod_{i=1}^{n} p(y_i \mid \theta, x_i) \tag{2}$$

The likelihood function $\ell(\theta)$, defined as the product of individual probability densities, is logarithmically transformed to enhance both parameter estimation efficiency and numerical stability, expressed as:

$$\mathcal{L}(\theta) = \log \ell(\theta) = \sum_{i=1}^{n} \log p(y_i \mid \theta, x_i) \tag{3}$$

In this study, the structural seismic response is assumed to follow a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ as the probabilistic model, with its probability density function (PDF) defined as:

$$p(y_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2} \right) \tag{4}$$

The log-likelihood function $\mathcal{L}(\theta)$ *under the Gaussian assumption* $\mathcal{N}(\mu, \sigma^2)$ *can be expressed as:*

$$\mathcal{L}(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu(x_i))^2 \tag{5}$$

In probabilistic ML prediction tasks, model performance is evaluated by maximizing the log-likelihood function $\mathcal{L}(\mu, \sigma^2)$. *To derive the MLE for the parameters* μ *and* σ^2 *of a Gaussian distribution, partial derivatives of the log-likelihood function are computed with respect to each parameter and set to zero, yielding the following analytical solutions:*

$$\frac{\partial \mathcal{L}(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) \Rightarrow \hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{6}$$

$$\frac{\partial \mathcal{L}(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2 \Rightarrow \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_{MLE})^2 \tag{7}$$

Similarly, if the structural seismic response data is assumed to follow a log-normal distribution $LN(\mu, \sigma^2)$, *the MLE for its parameters can be derived through analogous procedures.*