

Applications - Lecture 3

Statistics in Gaussian Models

Ingénierie Mathématique & Informatique

09/10/2026

Section	Notion	Used in
3.1	Standard Gaussian vectors, Cochran's theorem	App. 2
3.2	Joint law of (X_n, S_n^2) , Student distribution	App. 1, App. 2
3.3	OLS estimator $\hat{\beta}$, joint law of $(\hat{\beta}, \hat{\sigma}^2)$	App. 1

1. **Estimate & test** a machine-learning interatomic potential's coefficients by OLS.
2. **Estimate** a covariance structure and build a portfolio: sample covariance in Markowitz theory.

App. 1 - Fitting a Linear Interatomic Potential by OLS

Training a pair potential on DFT energies

Problem. I have n reference energies E_i (e.g. from DFT) computed at known interatomic distances r_i , $i = 1, \dots, n$, and wants to fit a linear potential with a confidence interval on its coefficients.

Model (Section 3.3). Writing the energy as a linear combination of two radial basis functions $\varphi_1(r) = r^{-12}$, $\varphi_2(r) = r^{-6}$, this is the Gaussian linear model $y_n = x_n\beta + \varepsilon_n$ with $y_i = E_i$, $x_i = (\varphi_1(r_i), \varphi_2(r_i))$, and $\varepsilon_n \sim \mathcal{N}_n(0, \sigma^2 I_n)$. The OLS estimator is

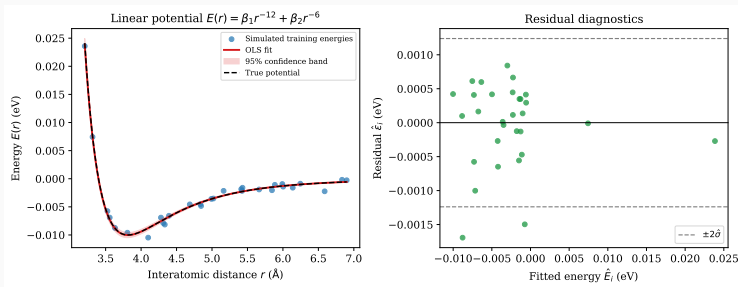
$$\hat{\beta} = (x_n^\top x_n)^{-1} x_n^\top y_n,$$

We know that $\hat{\beta}$ and $\hat{\sigma}^2 = \|y_n - \hat{y}_n\|^2 / (n - 2)$ are independent with

$$\hat{\beta} \sim \mathcal{N}_2(\beta, \sigma^2 (x_n^\top x_n)^{-1}), \quad \frac{(n - 2)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n - 2),$$

giving a $t(n - 2)$ pivot and confidence interval for each coefficient $\hat{\beta}_j$.

Fit and residual diagnostics



Left: $n = 30$ simulated training energies on a toy Lennard-Jones curve, OLS fit (red) with 95% confidence band, true potential (dashed). Right: residuals vs. fitted energies, $\pm 2\hat{\sigma}$ band.

- The residual plot checks $\text{Cov}[\varepsilon_n] = \sigma^2 I_n$, the validation step a fitted potential must pass before use.
- Lecture 6 revisits this model with a Bayesian prior for uncertainty quantification; Lecture 9 tests how many basis functions are actually needed.

App. 2 - Covariance Estimation and Portfolio Theory

Estimating a covariance structure

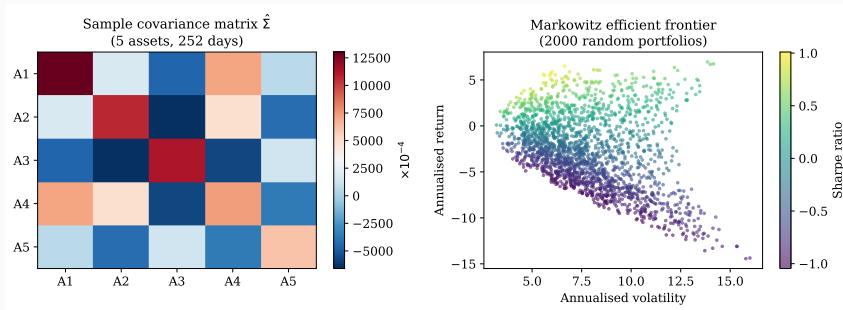
Problem. A portfolio manager holds p assets with daily returns $r_t = (r_{t,1}, \dots, r_{t,p})^\top$, $t = 1, \dots, n$ i.i.d., mean μ , covariance Σ . Each column $(r_{1,i}, \dots, r_{n,i})$ is a Gaussian sample.

Sample covariance. The unbiased estimator generalises S_n^2 :

$$\hat{\Sigma}_{ij} = \frac{1}{n-1} \sum_{t=1}^n (r_{t,i} - \bar{r}_i)(r_{t,j} - \bar{r}_j), \quad \hat{\Sigma}_{ii} = S_{n,i}^2.$$

Then $(n-1)S_{n,i}^2/\sigma_i^2 \sim \chi^2(n-1)$ gives a confidence interval for volatility σ_i^2 ; and $\sqrt{n}(\bar{r}_i - \mu_i)/S_{n,i} \sim t(n-1)$ tests if a return is significantly nonzero. **Markowitz** minimises $w^\top \hat{\Sigma} w$ ($\mathbf{1}^\top w = 1$) for a target return $w^\top \hat{\mu}$: the efficient frontier.

Reading the covariance matrix and the frontier



Left: $\hat{\Sigma}$ ($\times 10^4$), $p = 5$ synthetic assets, $n = 252$ trading days. Right: 2000 random portfolios (volatility, return), coloured by Sharpe ratio; upper-left boundary = efficient frontier.

- Warm off-diagonal entries mean assets move together, limiting diversification.
- No random portfolio reaches both low volatility and high return: $\hat{\Sigma}$ quantifies this risk-return tension.