Gaussian vectors

1. Let Σ be a symmetric positive definite matrix of $\mathbb{R}^{n\times n}$. Provide a solution to sample a Gaussian vector with covariance matrix Σ based on i.i.d. standard Gaussian variables.

2. Let ε be a random variable in $\{-1,1\}$ such that $\mathbb{P}(\varepsilon = 1) = 1/2$. If $(X,Y)^\top \sim \mathcal{N}(0, I_2)$ explain why the following vectors are or are not Gaussian vectors.

   (a) $(X, \varepsilon X)$.

   (b) $(X, \varepsilon Y)$.

   (c) $(X, \varepsilon X + Y)$.

   (d) $(X, X + \varepsilon Y)$.

3. Let X be a Gaussian vector in $\mathbb{R}^n$ with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\sigma^2 I_n$. Prove that the random variables $\bar{X}_n$ and $\widehat{\sigma}^2_n$ defined as

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \quad \text{and} \quad \widehat{\sigma}^2_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

   are independent.

Regression: prediction of a new observation

Consider the regression model given by

$$Y = X\beta_\star + \xi\,,$$

where $X \in \mathbb{R}^{n\times d}$ the $(\xi_i)_{1\leqslant i\leqslant n}$ are i.i.d. centered Gaussian random variables with variance $\sigma^2_\star$. Assume that $X^\top X$ has full rank and that $\beta_\star$ and $\sigma^2_\star$ are estimated by

$$\widehat{\beta}_n = (X^\top X)^{-1}X^\top Y \quad \text{and} \quad \widehat{\sigma}^2_n = \frac{\|Y - X\widehat{\beta}_n\|^2}{n-d}\,.$$

Let $x_\star \in \mathbb{R}^d$ and assume that its associated observation $Y_\star = x_\star^\top \beta_\star + \varepsilon_\star$ is predicted by $\widehat{Y}_\star = x_\star^\top \widehat{\beta}_n$.

1. Provide the expression of $\mathbb{E}[(\widehat{Y}_\star - x_\star^\top\beta_\star)^2]$.

2. Provide a confidence interval for $x_\star^\top \beta_\star$ with statistical significiance $1-\alpha$ for $\alpha \in (0,1)$.

Regression: linear estimators

Consider the regression model given, for all $1 \leqslant i \leqslant n$, by

$$Y_i = f^*(X_i) + \xi_i\,,$$

where for all $1 \leqslant i \leqslant n$, $X_i \in \mathsf{X}$, and the $(\xi_i)_{1\leqslant i\leqslant n}$ are i.i.d. centered Gaussian random variables with variance σ^2. In this exercise, f^* is estimated by a linear estimator of the form

$$\widehat{f}_n : x \mapsto \sum_{i=1}^n w_i(x) Y_i\,.$$

Prove that

$$\frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(\widehat{f}_n(X_i) - f^*(X_i))^2\right] = \frac{1}{n}\|Wf^*(X) - f^*(X)\|_2^2 + \frac{\sigma^2}{n}\mathrm{Trace}(W^\top W)\,,$$

where $W = (w_i(X_j))_{1\leqslant i,j\leqslant n}$ and $f^*(X) = (f^*(X_1), \ldots, f^*(X_n))^\top$.