ESILV ENGINEERING SCHOOL DE VINCI PARIS	**Advanced Statistics : Gaussian Linear Model**	**A4 ESILV** **Tutorial 2**

Exercise 1. Chi-Squared Distribution
Let X_1 and X_2 be independent random variables with distribution $\chi^2(m)$ and $\chi^2(n)$ respectively. Show that $X_1 + X_2$ follows a Khi-square distribution $\chi^2(m+n)$ too.

Exercise 2. Fisher Distribution
Let F be a random variable with a Fischer $\mathcal{F}(s,r)$ distribution. Show the following properties:

1. $\frac{1}{F}$ also has a Fischer $\mathcal{F}(s,r)$ distribution.
2. Let T a random variable with Student $\mathcal{T}(r)$ distribution. Then, prove that T^2 has a Fischer $\mathcal{F}(1,r)$ distribution.

Exercise 3. Empirical Mean and Variance
Let $X_1, X_2, \ldots, X_n$ be a Gaussian sample with mean μ and variance σ^2. We consider the empirical mean and variance defined by

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

The objective of this exercise is to determine the distribution of the unbiased empirical variance S_n^2 without relying on Cochran's Theorem.

1. Determine the distribution of
$$V = \frac{\sqrt{n}}{\sigma^2}(\overline{X}_n - \mu).$$
2. Assume that U and V are independent random variables. Find the distribution of
$$U = \frac{n-1}{\sigma^2} S_n^2.$$
3. Prove that U and V are independent. Then, state the result obtained in the previous question.
4. Determine the distribution of the statistic
$$T = \frac{V}{\sqrt{\frac{U}{n-1}}}.$$

Exercise 4. Ordinary Gaussian Model
Consider the following model:

$$Y_i = m + \sigma\varepsilon_i, \quad 1 \leq i \leq n,$$

where ε_i are i.i.d random variables with distribution $\mathcal{N}(0,1)$, the real parameter $m \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+^*$. We denote $\overline{Y}_n$ the empirical mean of random variable Y_i.

1. We assume that σ is a known parameter.

   (a) Let $\alpha \in (0,1)$. Give a symmetric confidence interval for m of level $1-\alpha$.

(b) Construct a test hypothesis of level α for $H_0 : m = m_0$ against $H_1 : m \neq m_0$.

2. Now, we assume that σ is unknown. We consider

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2.$$

(a) Specify the associated regression model and give the least square estimator of m.

(b) Remind the distribution of $\hat{\sigma}_n^2$ and give the distribution of $\sqrt{n}\frac{\overline{Y}_n - m}{\hat{\sigma}_n}$ by using Cochran's Theorem.

(c) Show that $\hat{\sigma}_n^2$ is an unbiased and consistent estimator of σ^2.

(d) Let $\alpha \in (0, 1)$. Give a confidence interval of level $1 - \alpha$ for σ^2.

(e) Construct a test of hypothesis for $H_0 : \sigma^2 = 3$ against $H_1 : \sigma^2 \neq 3$ of level α.

(f) Let $\alpha \in (0, 1)$. Give a confidence interval of level $1 - \alpha$ for m. Deduce a test for $H_0 : m = m_0$ against $H_1 : m \neq m_0$ of level α.

(g) Propose a test for $H_0 : m \geq m_0$ against $H_1 : m < m_0$ of level α. Calculate the p-value of the test for $n = 25$, $m_0 = 12.5$, $\overline{y} = 12$ and $\hat{\sigma}_n^2 = 1.69$? Can we reject H_0 at level 5% ? Indication: $F_{t_{24}}(-1.92) \simeq 0.03$.

Exercise 5. Simple Linear Gaussian Regression Model

Consider the following model:

$$Y_i = a + bt_i + \varepsilon_i, \quad 1 \leq i \leq n,$$

where ε_i are i.i.d random variables with distribution $\mathcal{N}(0, 1)$, the real numbers t_i are known for $1 \leq i \leq n$ and a, b, σ^2 are three real unknown parameters. We suppose $\sum_{i=1}^{n} t_i = 0$ and $\sum_{i=1}^{n} t_i^2 > 0$ and we denote by

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \;\; v_t = \frac{1}{n} \sum_{i=1}^{n} t_i^2, \;\; v_Y = \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \overline{Y}^2, \;\; \text{and} \;\; \rho = \frac{1}{n} \sum_{i=1}^{n} Y_i t_i.$$

1. Precise the identifiability conditions.

2. Calculate the least squares estimators $\hat{a}$, $\hat{b}$, $\hat{\sigma}^2$ of a, b and σ^2 as a function of Y, v_t, v_Y and ρ. What is their joint distribution?

3. Let $\alpha \in (0, 1)$. Give a confidence interval of level $1 - \alpha$ for each parameter a and b.

4. Construct a confidence ellipse ε of level $1 - \alpha$ for (a, b).

5. Give a confidence interval for $5a - 8b$, of level $1 - \alpha$.

6. Construct a hypothesis test for $H_0 : a = b$ against $H_1 : a \neq b$ of level α.

Exercise 6. Multiple Linear Regression Model

For $n \geq 4$, we consider the linear model $\mathbb{E}[Y] = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$ where $X_j \in \mathbb{R}^n$, with $j = 1, 2, 3$, are deterministic and known vectors. Furthermore we assume $\|X_j\|_{\mathbb{R}^n} = 1$ and for $i \neq j$, $\langle X_i, X_j \rangle_{\mathbb{R}^n} = 1/2$.

1. Express the model in the form $Y = X\beta + \varepsilon$ by specifying the assumptions and dimensions of the parameters.

2. For $q \in \mathbb{R}$, calculate the product of the square matrices A and B which are given by
    - $A = (A_{ij})_{1 \leq i,j \leq n}$ such that $A_{ii} = 1$ and for $i \neq j$ $A_{ij} = 1/2$.
    - $B = (B_{ij})_{1 \leq i,j \leq n}$ such that $B_{ii} = q$ and for $i \neq j$ $A_{ij} = -1$.
3. Calculate the least squares estimator $\hat{\beta}$ of β and deduce an unbiased estimator $\hat{\sigma}^2$ of σ^2. We assume for the following that the model is Gaussian, i.e., that ε is a centered Gaussian vector with covariance matrix $\sigma^2 \mathcal{I}_n$.
4. What is the joint distribution of $(\hat{\beta}, \hat{\sigma}^2)$?
5. Give a confidence interval for β_1 at level $1 - \alpha$.
6. Provide a hypothesis test of level α for $H_0 : \beta_2 = \beta_3 = 0$.