Gaussian vectors, Cochran's theorem, and applications to statistics

Gaussian variables are ubiquitous in probability and statistics. In fact, they can be encountered everywhere from thermodynamics to optics, to economics, to finance, and so on. In this problem, we introduce the concept of Gaussian vectors and introduce several classical distributions associated with the Gaussian one: Gamma, χ^2 and Student-t. The problem is guided by the goal of proving Cochran's theorem and that of presenting the most classical Student-t test: the location test for a Gaussian i.i.d. sample with unknown variance.

The Gamma function

Let us define the function Γ by

$$\Gamma : x \in \mathbb{R}_+^* \mapsto \int_0^{+\infty} t^{x-1} e^{-t} dt.$$

1. Prove that Γ is well defined and that $\forall x \in \mathbb{R}_+^*, \Gamma(x+1) = x\Gamma(x)$.

2. Prove that $\forall n \in \mathbb{N}^*, \Gamma(n) = (n-1)!$.

3. Prove that $\forall n \in \mathbb{N}, \Gamma\left(n + \frac{1}{2}\right) = \frac{(2n)!}{2^{2n} n!}\sqrt{\pi}$.

Classical distributions: Gaussian, Gamma, χ^2, and Student-t

Let $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$. We say that a real-valued random variable X follows a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ if there exists a random variable Z with probability density function

$$t \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

such that $X = \mu + \sigma Z$.

Let $\alpha, \lambda \in \mathbb{R}_+^*$. We say that a real-valued random variable X follows a Gamma distribution $\Gamma(\alpha, \lambda)$ if it has the following probability density function:

$$t \mapsto 1_{t>0} \frac{1}{\Gamma(\alpha)} \lambda^\alpha t^{\alpha-1} e^{-\lambda t}.$$

4. Let α, β, and λ be 3 positive real numbers. Prove that if X and Y are two independent random variables with respective distributions $\Gamma(\alpha, \lambda)$ and $\Gamma(\beta, \lambda)$ then $X + Y$ follows a $\Gamma(\alpha + \beta, \lambda)$ distribution.

5. Let X be a $\mathcal{N}(0, 1)$ random variable. What is the distribution of X^2?

6. Let $X_1, \ldots, X_n$ be n i.i.d $\mathcal{N}(0, 1)$ random variables. Prove that $X_1^2 + \cdots + X_n^2$ follows a $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ distribution.

Remark-Definition: The distribution $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ is called χ_n^2 (and reads "kai" square with n degrees of freedom).

7. Let Z be a $\mathcal{N}(0,1)$ random variable. Let X be a random variable with a χ_n^2 distribution (for some $n \in \mathbb{N}^*$). If X and Z are independent, prove that $\frac{Z}{\sqrt{\frac{X}{n}}}$ has the following probability density function:

$$f_n : t \mapsto \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

Remark-Definition: The above probability density function defines a distribution called Student-t distribution with n degrees of freedom).

8. Prove that $(f_n)_{n\in\mathbb{N}^*}$ converges pointwise towards the probability density function of a $\mathcal{N}(0,1)$ random variable.

Characteristic function

For what follows, we recall the following:

- For any $\mathbb{R}^n$-valued random variable X, the characteristic function of X is the function

$$\phi_X : \xi \in \mathbb{R}^d \mapsto \mathbb{E}[\exp(i\xi' X)],$$

  where $'$ is the transposition operator.

  Remark: If X is $\mathbb{R}$-valued, we rather write $\phi_X : t \in \mathbb{R} \mapsto \mathbb{E}[\exp(itX)]$.

- If X and Y are two $\mathbb{R}^n$-valued random variables, $\phi_X = \phi_Y$ implies that X and Y have the same distribution.

Let Z be a Gaussian $\mathcal{N}(0,1)$ random variable.

9. Prove that ϕ_Z is a function of class C^1 and that $\forall t \in \mathbb{R}, \phi_Z'(t) = -t\phi_Z(t)$.

10. Deduce that $\forall t \in \mathbb{R}, \phi_Z(t) = \exp\left(-\frac{t^2}{2}\right)$.

Let X be a Gaussian $\mathcal{N}(\mu, \sigma^2)$ random variable.

11. Prove that $\forall t \in \mathbb{R}, \phi_X(t) = \exp\left(it\mu - \frac{1}{2}\sigma^2 t^2\right)$.

Let X and Y be two independent Gaussian random variables.

12. Prove that any linear combination of X and Y is Gaussian.

Gaussian vectors

Let $n \in \mathbb{N}^*$ and let us consider a $\mathbb{R}^n$-valued random variable. We say that X is a Gaussian vector if and only if $\forall \xi \in \mathbb{R}^n$, $\xi' X$ is a Gaussian variable.

If $X = (X_1, \ldots, X_n)$ is a $\mathbb{R}^n$-valued Gaussian vector, we denote by μ_X the vector $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_n])'$ and by Σ_X the matrix $(\mathrm{Cov}(X_i, X_j))_{1\leq i,j\leq n}$.

13. Prove that if $X = (X_1, \ldots, X_n)$ is a $\mathbb{R}^n$-valued Gaussian vector, then $\forall i \in \{1, \ldots, n\}$, X_i is Gaussian. Is the reciprocal true?

Hint: For the reciprocal, consider X and ϵX where X a Gaussian $\mathcal{N}(0,1)$ random variable and ϵ a random variable with $\mathbb{P}(\epsilon=-1)=\mathbb{P}(\epsilon=1)=\frac{1}{2}$.

14. Let $X_1,\dots,X_n$ be n independent $\mathbb{R}$-valued Gaussian random variable. Prove that $X=(X_1,\dots,X_n)$ is a $\mathbb{R}^n$-valued Gaussian vector.

Let X be a $\mathbb{R}^n$-valued Gaussian vector.

15. Let $\xi\in\mathbb{R}^n$. Show that $\mathbb{E}[\xi' X]=\xi'\mu_X$ and $\mathbb{V}[\xi' X]=\xi'\Sigma_X\xi$.

16. Deduce that $\phi_X:\xi\in\mathbb{R}^n\mapsto\exp\left(i\xi'\mu_X-\frac{1}{2}\xi'\Sigma_X\xi\right)$.

17. Prove that Σ_X is a positive semidefinite symmetric matrix.

18. Let $m\in\mathbb{N}^*$. Prove that for any matrix $A\in\mathbb{M}_{m,n}$, AX is a $\mathbb{R}^m$-valued Gaussian vector. Prove that $\mu_{AX}=A\mu_X$ and $\Sigma_{AX}=A\Sigma_X A'$.

Let $p\in\mathbb{N}\setminus\{0,1\}$. Let $n_1,\dots,n_p\in\mathbb{N}^*$ and let $n=n_1+\ldots+n_p$. Assume that there exists p positive semidefinite symmetric matrices $\Sigma_1,\dots,\Sigma_p$ of respective size $n_1,\dots,n_p$ such that

$$\Sigma_X=\begin{pmatrix}\Sigma_1 & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \Sigma_p\end{pmatrix}.$$

19. Prove that the random variables
$$(X_1,\dots,X_{n_1}),(X_{n_1+1},\dots,X_{n_1+n_2}),\dots,(X_{n_1+\ldots+n_{p-1}+1},\dots,X_n)$$
are independent Gaussian vectors.

Cochran' theorem

Let $n\in\mathbb{N}^*$ and let $X_1,\dots,X_n$ be n independent $\mathcal{N}(0,1)$ Gaussian random variables.

Let $p\in\mathbb{N}^*$ and let $F_1,\dots,F_p$ be p vector subspaces of $\mathbb{R}^n$ with $\forall i\neq j, F_i\perp F_j$.

Let us denote by $P_1,\dots,P_p$ the matrices (in the canonical basis) of the orthogonal projections on $F_1,\dots,F_p$ respectively.

20. Prove that $P_1X,\ \dots,\ P_pX$ are independent Gaussian vectors with $\forall i\in\{1,\dots,p\},\mu_{P_iX}=0$ and $\Sigma_{P_iX}=P_i$.
Hint: consider AX where $A=\begin{pmatrix}P_1 & \cdots & P_p\end{pmatrix}'$.

21. Prove that $\forall i\in\{1,\dots,p\}$, $\|P_iX\|^2$ follows a $\chi^2_{\dim(F_i)}$ distribution.
Hint: Diagonalize P_i.

Application to statistics

Let $n\in\mathbb{N}^*$ and let $X_1,\dots,X_n$ be n independent $\mathcal{N}(\mu,\sigma^2)$ Gaussian random variables. Let us define $X=(X_1,\dots,X_n)'$.

Let us consider

$$\overline{X}_n = \frac{X_1 + \ldots + X_n}{n} \quad \text{and} \quad \widehat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}.$$

Let $u = (1, \ldots, 1)' \in \mathbb{R}^N$. Let $F = \text{span}(u)$ and let P_F and $P_{F^\perp}$ be the matrices (in the canonical basis) of the orthogonal projectors on F and $F^\perp$ respectively.

22. Prove that $P_F X = \overline{X}_n u$.

23. Deduce that $(n-1)\widehat{\sigma}_n^2 = \|P_{F^\perp} X\|^2$.

24. Deduce that $\frac{(n-1)\widehat{\sigma}_n^2}{\sigma^2}$ follows a χ^2_{n-1} distribution. Deduce the value of $\mathbb{E}[\widehat{\sigma}_n^2]$.

25. Prove that $\sqrt{n}\frac{\overline{X}_n - \mu}{\widehat{\sigma}_n}$ follows a Student-t distribution with $n-1$ degrees of freedom.

26. How can we exploit the previous question to build a location test where $\mathrm{H}_0 : \mu = 0$ and $\mathrm{H}_1 : \mu \neq 0$?