Properties of the Multivariate Normal Distribution

Recall the following definition.

Definition 1 *We say that* $\mathbf{Y}$ *is a* p*-dimensional standard normal vector if its components are independent standard normal variables. Let* $\mathbf{A}$ *be a* $p \times p$ *regular real matrix and* $\mathbf{m} \in \mathbb{R}^p$ *be a vector. Then the linear transformation* $\mathbf{X} = \mathbf{AY} + \mathbf{m}$ *defines a* p*-dimensional normal random vector.*

First we derive the p.d.f. of a multivariate normal (Gaussian) random vector.

- The random vector $\mathbf{Y}$ has *p-dimensional standard normal distribution*, if its components are i.i.d. standard Gaussian variables. Therefore, the p.d.f. of $\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ is

$$g(\mathbf{y}) = \prod_{i=1}^{p} \phi(y_i) = \frac{1}{\sqrt{2\pi}^p} e^{-(\sum_{i=1}^p y_i^2)/2} = \frac{1}{(2\pi)^{p/2}} e^{-\|\mathbf{y}\|^2/2}, \quad \mathbf{y} \in \mathbb{R}^p$$

  where ϕ is the standard Gaussian density and $\mathbf{y} = (y_1, \dots, y_p)^T$.

- The random vector $\mathbf{X} = \mathbf{AY} + \mathbf{m}$ has *p-dimensional normal distribution*, where $\mathbf{A}$ is a $p \times p$ regular matrix and $\mathbf{m} \in \mathbb{R}^p$.

  By the linearity of the expectation and the bilinearity of the covariance,

$$\mathbb{E}(\mathbf{X}) = \mathbf{A}\mathbb{E}\mathbf{Y} + \mathbf{m} = \mathbf{m}$$

  and the covariance matrix of $\mathbf{X}$ is

$$\mathbf{C} = \mathrm{Var}\mathbf{X} = \mathbb{E}[(\mathbf{X}-\mathbf{m})(\mathbf{X}-\mathbf{m})^T] = \mathbb{E}[(\mathbf{AY})(\mathbf{AY})^T] = \mathbb{E}(\mathbf{AYY}^T\mathbf{A}^T) = \mathbf{A}\mathbb{E}(\mathbf{YY}^T)\mathbf{A}^T = \mathbf{AA}^T.$$

  Notation: $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$, where $p, \mathbf{m}, \mathbf{C}$ are parameters. If $\mathbf{A}$, and equivalently, $\mathbf{C}$ is invertible, the p.d.f. $f(\mathbf{x})$ of $\mathbf{X}$ is derived by the transformation formula applied with the one-to-one $\mathbb{R}^p \to \mathbb{R}^p$ transformations

$$\mathbf{x} = \mathbf{Ay} + \mathbf{m}, \qquad \mathbf{y} = \mathbf{A}^{-1}(\mathbf{x} - \mathbf{m}).$$

  Namely,

$$\begin{aligned} f(\mathbf{x}) &= |\det(\mathbf{A}^{-1})| \cdot g(\mathbf{A}^{-1}(\mathbf{x} - \mathbf{m})) = \frac{1}{|\det(\mathbf{A})|} \cdot \frac{1}{(2\pi)^{p/2}} e^{-\|\mathbf{A}^{-1}(\mathbf{x}-\mathbf{m})\|^2/2} \\ &= \frac{1}{(2\pi)^{p/2}|\det(\mathbf{A})|} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T(\mathbf{A}^{-1})^T\mathbf{A}^{-1}(\mathbf{x}-\mathbf{m})} \\ &= \frac{1}{(2\pi)^{p/2}|\mathbf{C}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})}, \quad \mathbf{x} \in \mathbb{R}^p. \end{aligned}$$

  Here $|\det(\mathbf{A}^{-1})| = 1/|\det(\mathbf{A})|$ is the inverse Jacobian. We also used that $|\det(\mathbf{A})| = |\mathbf{C}|^{1/2}$ and $(\mathbf{A}^{-1})^T\mathbf{A}^{-1} = (\mathbf{A}^T)^{-1}\mathbf{A}^{-1} = (\mathbf{AA}^T)^{-1} = \mathbf{C}^{-1}$. Note that $\mathbf{C}$ is positive definite, therefore its determinant is positive, and we use the $|\mathbf{C}| = \det(\mathbf{C})$ notation.

We remark that a linear transformation with a singular or rectangular $\mathbf{A}$ would result in a degenerated p-variate normal distribution which, in fact, is realized in a lower, namely, rank($\mathbf{A}$)-dimensional subspace. Hence, we will only deal with regular $\mathbf{A}$ and $\mathbf{C}$. Note that rank($\mathbf{C}$) = rank($\mathbf{A}$).

Proposition 1 *The level surfaces (contours of equal density) of f are ellipsoids, which are spheres if and only if the components of* $\mathbf{X}$ *are independent with equal variances.*

Proof: Obviously, $\mathbf{f}(\mathbf{x}) > 0$ and it attains its maximum at $\mathbf{m}$. Let C be a constant with $0 < C \leq f(\mathbf{m})$. Then $f(\mathbf{x}) = C$ is equaivalent to

$$(\mathbf{x}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{m}) = \mathbf{z}^T\mathbf{\Lambda}^{-1}\mathbf{z} = \sum_{i=1}^{p}\frac{1}{\lambda_i}z_i^2 = \sum_{i=1}^{p}\frac{z_i^2}{\sqrt{\lambda_i}^2} = c,$$

where $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ and $\mathbf{C}^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^T$ is SD, and $\mathbf{z} = \mathbf{U}^T(\mathbf{x}-\mathbf{m})$ is the principal axes coordinate transformation. Further, $c > 0$ is a transformation of the constant C. The above equation is equivalent to

$$\sum_{i=1}^{p}\frac{z_i^2}{\sqrt{c\lambda_i}^2} = 1,$$

which is the equation of an ellipsoid with half-axes proportional to $\sqrt{\lambda_i}$'s in the new coordinate system. We have circles if and only if

$$\lambda_1 = \cdots = \lambda_p = \lambda,$$

i.e.,

$$\mathbf{C} = \mathbf{U}(\lambda\mathbf{I}_p)\mathbf{U}^T = \lambda\mathbf{I}_p$$

which means that the components are independent with equal variances λ. Note that if the components are independent, but not of equal variances, we have ellipsoids with axes parallel to the coordinate axes.

Conversely, given the parameters $\mathbf{m}$ and $\mathbf{C}$, $\mathbf{X}$ can be transformed into a p-dimensional standard normal vector $\mathbf{Y}$ in the following way. $\mathbf{C}$, being a Gram-matrix, can be (not uniquely) decomposed as $\mathbf{A}\mathbf{A}^T$ with a regular $p \times p$ matrix $\mathbf{A}$. For example, $\mathbf{A} = \mathbf{C}^{1/2}$ will do, but $\mathbf{AQ}$ is also convenient with any $p \times p$ orthogonal matrix $\mathbf{Q}$. Then the formula

$$\mathbf{Y} = \mathbf{A}^{-1}(\mathbf{X}-\mathbf{m})$$

defines a p-dimensional standard normal vector. Note that using $\mathbf{AQ}$ instead of $\mathbf{A}$ will result in an orthogonal rotation of $\mathbf{Y}$ which has the same distribution as $\mathbf{Y}$.

Proposition 2 *The components of* $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m},\mathbf{C})$ *are (completely) independent if and only if* $\mathbf{C}$ *is diagonal.*

Proposition 3 *If the covariance matrix* $\mathbf{C}$ *of* $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m},\mathbf{C})$ *is positive definite, then*

$$(\mathbf{X}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{X}-\mathbf{m}) \sim \chi^2(p).$$

Proof: With the above back transformation

$$(\mathbf{X}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{X}-\mathbf{m}) = (\mathbf{X}-\mathbf{m})^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{X}-\mathbf{m}) =$$

$$= (\mathbf{X}-\mathbf{m})^T(\mathbf{A}^{-1})^T\mathbf{A}^{-1}(\mathbf{X}-\mathbf{m}) = \mathbf{Y}^T\mathbf{Y} = \sum_{i=1}^{p}Y_i^2$$

which finishes the proof by the definition of the χ^2 distribution, since Y_i's are i.i.d. standard Gaussians.

Note that if $\text{rank}(\mathbf{C}) = r < p$, then

$$(\mathbf{X} - \mathbf{m})^T \mathbf{C}^+ (\mathbf{X} - \mathbf{m}) \sim \chi^2(r),$$

where $\mathbf{C}^+$ is the Moore–Penrose inverse of the singular matrix $\mathbf{C}$ (see Lesson 1).

Proposition 4 *The random vector* $\mathbf{X}$ *has multivariate normal distribution if and only if any linear combination of its components has one-dimensional normal distribution.*

This follows easily by characteristic functions (see the Multivar. Stat. material). Proposition 4 implies that all the marginal distributions of the multivariate normal distribution are multivariate normal of appropriate dimension and parameters.

As for the conditional distributions, in case of the multivariate normal distribution, the conditional expectations are linear functions of the subset of variables in the condition. More precisely, the following proposition can be proved.

Proposition 5 *Let* $(\mathbf{X}^T, \mathbf{Y}^T)^T \sim \mathcal{N}_{p+q}(\mathbf{m}, \mathbf{C})$ *be a random vector, where the expectation* $\mathbf{m}$ *and the covariance matrix* $\mathbf{C}$ *are partitioned (with block sizes* p *and* q*) in the following way:*

$$\mathbf{m} = \begin{pmatrix} \mathbf{m_X} \\ \mathbf{m_Y} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{C_{XX}} & \mathbf{C_{XY}} \\ \mathbf{C_{YX}} & \mathbf{C_{YY}} \end{pmatrix}.$$

Here $\mathbf{C_{XX}}$, $\mathbf{C_{YY}}$ *are covariance matrices of* $\mathbf{X}$ *and* $\mathbf{Y}$*, whereas* $\mathbf{C_{YX}} = \mathbf{C_{XY}}^T$ *is the cross-covariance matrix. Assume that* $\mathbf{C_{XX}}$*,* $\mathbf{C_{YY}}$ *and* $\mathbf{C}$ *are regular. Then the conditional distribution of the random vector* $\mathbf{Y}$ *conditioned on* $\mathbf{X}$ *is* $\mathcal{N}_q(\mathbf{C}_{\mathbf{Y}X}\mathbf{C}_{\mathbf{XX}}^{-1}(\mathbf{X} - \mathbf{m_X}) + \mathbf{m_Y}, \mathbf{C_{YY|X}})$ *distribution, where*

$$\mathbf{C_{YY|X}} = \mathbf{C_{YY}} - \mathbf{C_{YX}}\mathbf{C_{XX}}^{-1}\mathbf{C_{XY}}.$$

The conditional expectation of $\mathbf{Y}$ conditioned on $\mathbf{X}$ is the expectation of the above conditional distribution:

$$\mathbb{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{C_{YX}}\mathbf{C_{XX}}^{-1}(\mathbf{X} - \mathbf{m_X}) + \mathbf{m_Y}$$

which is a linear function of the coordinates of $\mathbf{X}$. In the $p = q = 1$ case, it is called *regression line*, while in the $q = 1$, $p > 1$ case, *regression plane*. We will not deal with the $q > 1$, $p > 1$ case, which is the topic of the Partial Least Squares Regression. Summarizing, in case of multidimensional Gaussian distribution the regression functions are linear functions of the variables in the condition, which fact has important consequences in multivariate statistical analysis.

Note that multivariate normality can often be assumed due to the forthcoming *Multidimensional Central Limit Theorem.* As a preparation to this theorem, we consider independent sums.

Proposition 6 (Multidimensional Steiner Theorem) *Let* $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ *be given vectors,* $\overline{\mathbf{x}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k$*, and* $\mathbf{v} \in \mathbb{R}^p$ *be an arbitrary vector. Then*

$$\sum_{k=1}^{n} (\mathbf{x}_k - \mathbf{v})(\mathbf{x}_k - \mathbf{v})^T = \sum_{k=1}^{n} (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T + n(\bar{\mathbf{x}} - \mathbf{v})(\bar{\mathbf{x}} - \mathbf{v})^T.$$

Especially, with $\mathbf{v} = \mathbf{0}$ *we get*

$$\sum_{k=1}^{n} (\mathbf{x}_k - \bar{\mathbf{x}})(\mathbf{x}_k - \bar{\mathbf{x}})^T = \sum_{k=1}^{n} \mathbf{x}_k \mathbf{x}_k^T - n\bar{\mathbf{x}}\bar{\mathbf{x}}^T.$$

Proposition 7 *Let* $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ *be i.i.d. random vectors with expectation vector* $\mathbf{m}$ *and covariance matrix* $\mathbf{C}$*. Then expectation vector and covariance matrix of* $\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i$*:*

$$\mathbb{E}\overline{\mathbf{X}} = \mathbf{m}, \quad \mathrm{Var}\overline{\mathbf{X}} = \frac{1}{n}\mathbf{C}.$$

Theorem 1 (Multidimensional Central Limit Theorem) *Let* $\mathbf{X}_1, \mathbf{X}_2, \dots$ *be i.i.d. random vectors. Assume that the common expectation vector* $\mathbf{m}$ *and the covariance matrix* $\mathbf{C}$ *of* $\mathbf{X}_i$*'s exist, and* $\mathbf{C}$ *is regular. Then the sequence of the standardized partial sums*

$$\frac{1}{\sqrt{n}} \mathbf{C}^{-1/2} \left(\sum_{i=1}^{n} \mathbf{X}_i - n\mathbf{m} \right)$$

converges (in distribution) to the p*-dimensional standard normal distribution as* $n \to \infty$*. Equivalently,*

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \mathbf{X}_i - n\mathbf{m} \right) \to \mathcal{N}_p(\mathbf{0}, \mathbf{C}).$$

The convergence is understood in distribution, which means the convergence of the cumulative multivariate distribution functions. (Not proved here, but can be proved by characteristic functions.)

Definition 2 *The empirical and corrected empirical covariance matrices based on the* $\mathbf{X}_1, \dots, \mathbf{X}_n$ *i.i.d. sample are*

$$\hat{\mathbf{C}} = \frac{1}{n}\mathbf{S} \quad \text{and} \quad \hat{\mathbf{C}}^* = \frac{1}{n-1}\mathbf{S},$$

where

$$\mathbf{S} = \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^T - n\bar{\mathbf{X}}\bar{\mathbf{X}}^T.$$

Proposition 8 *The corrected empirical covariance matrix is unbiased, while the empirical one is asymptotically unbiased estimator of the true covariance matrix* $\mathbf{C}$*.*

Proof:

$$\mathbb{E}\mathbf{S} = \sum_{i=1}^{n} \mathbb{E}(\mathbf{X}_i\mathbf{X}_i^T) - n\mathbb{E}(\bar{\mathbf{X}}\bar{\mathbf{X}}^T) = n(\mathbf{C} - \mathbf{m}\mathbf{m}^T) - n(\frac{1}{n}\mathbf{C} - \mathbf{m}\mathbf{m}^T) = (n-1)\mathbf{C},$$

which finishes the proof.

Prove Propositions 4,5,6,7!

Characteristic functions help us to prove Proposition 4.

Proposition 9 *The characteristic function of* $\mathbf{Y} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}_p)$ *is*

$$\psi_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{Y}^T\mathbf{t}}) = e^{-\|\mathbf{t}\|^2/2}, \quad \mathbf{t} \in \mathbb{R}^p.$$

Proposition 10 *The characteristic function of* $\mathbf{X} \sim \mathcal{N}_p(\mathbf{m}, \mathbb{C})$ *is*

$$\psi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{X}^T\mathbf{t}}) = e^{i\mathbf{m}^T\mathbf{t} - \frac{1}{2}\mathbf{t}^T\mathbb{C}\mathbf{t}}, \quad \mathbf{t} \in \mathbb{R}^p,$$

where i is the imaginary unit.

Proposition 11 *If X and Y are independent and $X+Y$ is normally distributed, then X és Y are also normally distributed.*

(Not proved, but can be proved by characteristic functions.)

Lemma 1 *Let* $\mathbf{X} = (X_1, \dots, X_k)^T \sim \mathcal{N}_k(\mathbf{0}, \mathbf{C})$ *with* $\mathbf{C}$ *positive semidefinite. Then* $\sum_{i=1}^k X_i^2$ *can be decomposed as* $\sum_{i=1}^k \lambda_i Y_i^2$*, where* $\mathbf{Y} = (Y_1, \dots, Y_k)^T \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}_k)$ *and the nonnegative real numbers* $\lambda_1, \dots, \lambda_k$ *are eigenvalues of* $\mathbf{C}$.

Proof. Let the SD of the covariance matrix of $\mathbf{X}$ be $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_k)$. Then $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{Y}$ where $\mathbf{Y}$ is k-dimensional standard normal.

$$\sum_{i=1}^{k} X_i^2 = \|\mathbf{X}\|^2 = \|\mathbf{U}^T\mathbf{X}\|^2 = \|\mathbf{\Lambda}^{1/2}\mathbf{Y}\|^2 = \sum_{i=1}^{k} \lambda_i Y_i^2,$$

as the orthogonal transformation $\mathbf{U}^T$ keeps the norm. This was to be proven.

Based on this, we are able to find the asymptotic distribution of the well-known χ^2 statistic.

Revisiting the χ^2-test. Let $A_1, \dots, A_k$ be a complete set of mutually exclusive events. Check

$$H_0 : \mathbf{P}(A_i) = p_i \quad (i = 1, \dots, k).$$

Denote by $\nu_1, \dots, \nu_k$ the frequencies of $A_1, \dots, A_k$ in n independent trials ($\sum_{i=1}^k \nu_i = n$). Then under the zero-hypothesis

$$\underline{\nu} = (\nu_1, \dots, \nu_k)^T \sim \mathcal{P}oly_n(p_1, \dots, p_k).$$

(Recall that it is a deformed k-dimensional distribution concentrated on a $(k-1)$-dimensional hyperplane of $\mathbb{R}^k$ because of the linear relation $\nu_1 + \dots + \nu_k = n$ between its components.)

Theorem 2 *If $\underline{\nu} = (\nu_1, \dots, \nu_k)^T$ follows polynomial distribution with parameters n and $p_1, \dots, p_k$ $\quad (p_i > 0,\ \sum_{i=1}^k p_i = 1)$, then*

$$\sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i} \to \chi^2(k-1)$$

in distribution as $n \to \infty$.

Proof. First find the expectation vector and covariance matrix of the polynomially distributed random vector $\underline{\nu} = (\nu_1, \dots, \nu_k)^T$.

The k-dimensional indicator of $A_1, \dots, A_k$ is a random vector $\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_k)^T$, the components of which corresponding to the actual random event is 1, the others are 0s. If $\varepsilon_k^{(i)}$ denotes the indicator of the ith experiment, then $\underline{\nu} = \sum_{i=1}^n \underline{\varepsilon}^{(i)}$.

For the components of the expectation vector of $\underline{\varepsilon}$:

$$\mathbb{E}(\varepsilon_j) = \mathbf{P}(A_j) = p_j, \qquad j = 1, \dots, k.$$

and the diagonal entries of its covariance matrix are

$$c_{jj} = \mathbf{D}^2(\varepsilon_j) = \mathbb{E}(\varepsilon_j^2) - \mathbb{E}^2(\varepsilon_j) = p_j - p_j^2 = p_j(1-p_j), \qquad j = 1, \dots, k,$$

while the off-diagonal entries are

$$c_{lj} = c_{jl} = \mathrm{Cov}(\varepsilon_j, \varepsilon_l) = \mathbb{E}(\varepsilon_j \varepsilon_l) - \mathbb{E}(\varepsilon_j) \cdot \mathbb{E}(\varepsilon_l) = 0 - p_j p_l = -p_j p_l, \quad 1 \le j < l \le k.$$

Therefore, the covariance matrix of $\underline{\varepsilon}$ is

$$\mathbf{C} = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & \dots & -p_2p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_k & -p_2p_k & \dots & p_k(1-p_k) \end{pmatrix} = \mathbf{P}\mathbf{I}_k - \mathbf{p}\mathbf{p}^T, \qquad (1)$$

where $\mathbf{P} = \mathrm{diag}(p_1, \dots, p_k)$ és $\mathbf{p} = (p_1, \dots, p_k)^T$. So, $\mathrm{Var}(\underline{\nu}) = n\mathbf{C}$. Obviously, $\underline{\nu}$ and $\mathbf{C}$ are singular (their row sums are 0s).

Apply the Multivariate Central Limit Theorem to the i.i.d. $\mathbf{P}^{-1/2}\underline{\varepsilon}^{(1)}, \dots, \mathbf{P}^{-1/2}\underline{\varepsilon}^{(n)}$ random vectors, the sum of which is $\mathbf{P}^{-1/2}\underline{\nu}$. So, as $n \to \infty$, $\frac{1}{\sqrt{n}}\mathbf{P}^{-1/2}(\underline{\nu} - \mathbb{E}(\underline{\nu}))$ tends (in distribution) to a random vector $\mathbf{X} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{P}^{-1/2}\mathbf{C}\mathbf{P}^{-1/2})$.

But by the decomposition (1) of the matrix $\mathbf{C}$:

$$\mathbf{P}^{-1/2}\mathbf{C}\mathbf{P}^{-1/2} = \mathbf{I}_k - \mathbf{Q},$$

where $\mathbf{Q}$ is the dyad of the unit vector $\mathbf{q} = (\sqrt{p_1}, \dots, \sqrt{p_k})^T$. Consequently, $\mathbf{Q}$ defines a projection onto the direction a$\mathbf{q}$, and $\mathbf{I}_k - \mathbf{Q}$ a projection onto the orthogonal complementary subspace of $\mathbf{q}$. The eigenvalues of this projection of rank $(k-1)$ are $\lambda_1 = \dots = \lambda_{k-1} = 1$ és $\lambda_k = 0$.

Apply Lemma 1 to $\mathbf{X}$: the distribution of $\sum_{i=1}^k X_i^2$ is the same as that of $\sum_{i=1}^k \lambda_i Y_i^2$, where $Y_i$are i.i.d. standard normals, and λ_is are the eigenvalues of $\mathbf{I}_k - \mathbf{Q}$. So $\sum_{i=1}^k X_i^2 \sim \chi^2(k-1)$.

By a general probability theorem it follows that if $\frac{1}{\sqrt{n}}\mathbf{P}^{-1/2}(\underline{\nu}-\mathbb{E}(\underline{\nu}))$ tends in distribution to $\mathbf{X}$ $(n \to \infty)$, then the sum of the squares of its coordinates (what is our χ^2-statistic) tends in distribution to the sum of the squares of $\mathbf{X}$, which is the $\chi^2(k-1)$-distribution. This finishes the proof.