Regression: Theory and Practice

Gaussian vectors

Karim Lounici

Random vectors

- Let $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix} \in \mathbb{R}^d$ be a random vector such that $\mathbb{E}[X_j^2] < \infty$ for all $1 \leq j \leq d$.
- For all $1 \leq i, j \leq d$, the covariance between X_i and X_j is

$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j].$$

- Remarks:
  - Symmetry: $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$
  - If X_i, X_j are independent then $\mathrm{Cov}(X_i, X_j) = 0$
  - The converse is false in general.

Random vectors

- Expectation vector:

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \mathbb{E}[X_2] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix}$$

- Covariance operator

$$\Sigma_X = \mathrm{Cov}(X) := \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_d) \\ \mathrm{Cov}(X_1, X_2) & \mathrm{Var}(X_1) & \cdots & \mathrm{Cov}(X_2, X_d) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_d, X_1) & \mathrm{Cov}(X_d, X_2) & \cdots & \mathrm{Var}(X_d) \end{pmatrix}$$

Theorem

Σ_X *is symmetric positive semi-definite* ($v^\top Cov(X) v \geq 0, \forall v \in \mathbb{R}^d$)

Proof.

By definition, $v^\top Cov(X) v = \mathrm{Var}(v^\top X) \geq 0$ $\square$

Random vectors

Theorem (Affine mapping)

Let $\alpha \in \mathbb{R}^{\cancel{d}\,k}$ and $A \in \mathbb{R}^{k\times d}$, then $Y = \alpha + A\,X$ is a random vector in $\mathbb{R}^k$ with

$$\mathbb{E}[Y] = \alpha + A\,\mathbb{E}[X], \quad \textit{and} \quad \Sigma_Y = A\Sigma_X A^\top.$$

Proof as an exercise.

$X \in \mathbb{R}^d, \; Y \in \mathbb{R}^k$

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY^T] - \mathbb{E}[X] \cdot (\mathbb{E}[Y])^T$$

Random vectors

Theorem (Affine mapping)

Let $\alpha \in \mathbb{R}^d$ and $A \in \mathbb{R}^{k \times d}$, then $Y = \alpha + A\,X$ is a random vector in $\mathbb{R}^k$ with

$$\mathbb{E}[Y] = \alpha + A\,\mathbb{E}[X], \quad \textit{and} \quad \Sigma_Y = A \Sigma_X A^\top.$$

Proof as an exercise.

Definition (characteristic function)

Let $X \in \mathbb{R}^d$ be a random vector. Its characteristic function $\Phi_X : \mathbb{R}^d \to \mathbb{C}$ is

$$\Phi_X(t) = \mathbb{E}[e^{i\,\langle t, X \rangle}] = \mathbb{E}[e^{i\,\sum_{j=1}^d t_j X_j}].$$

- X and $\tilde{X}$ random vectors in $\mathbb{R}^d$ are independent iff $\phi_{X+\tilde{X}}(t) = \phi_X(t)\phi_{\tilde{X}}(t)$, for any $t \in \mathbb{R}^d$.

Gaussian Vectors

Definition

a random variable $z \in \mathbb{R}$ is Gaussian iff its characteristic function is

$$\Phi_z(t) = e^{im - \frac{1}{2}t^2\sigma^2}, \quad \forall t \in \mathbb{R}.$$

z is a degenerate Gaussian if $\sigma = 0$.

Definition

a random vector $X \in \mathbb{R}^d$ is a Gaussian vector iff $a^\top X$ is a Gaussian random variable for any $a \in \mathbb{R}^d$. normal

- If $\mathbb{E}[X] = 0$ and $\mathrm{Cov}(X) = I_d$, then we say that the Gaussian vector X is standard.
- If $X \in \mathbb{R}^d$ is a Gaussian vector, then its components are Gaussian rv's.

$a = e_j = (0, \ldots, 0, 1, 0, \ldots, 0)^\top \in \mathbb{R}^d$

$e_j^\top X = X_j$

Gaussian Vectors

Definition

a random variable $z \in \mathbb{R}$ is Gaussian iff its characteristic function is

$$\Phi_z(t) = e^{im - \frac{1}{2}t^2\sigma^2}, \quad \forall t \in \mathbb{R}.$$

z is a degenerate Gaussian if $\sigma = 0$.

Definition

a random vector $X \in \mathbb{R}^d$ is a Gaussian vector iff $a^\top X$ is a Gaussian random variable for any $a \in \mathbb{R}^d$.

- If $\mathbb{E}[X] = 0$ and $\mathrm{Cov}(X) = I_d$, then we say that the Gaussian vector X is standard.
- If $X \in \mathbb{R}^d$ is a Gaussian vector, then its components are Gaussian rv's.

Question: If all the components of a Gaussian vector X are Gaussian rv's, is X a Gaussian vector?

Gaussian Vectors

Theorem

The following assertions are equivalent:
- $X \in \mathbb{R}^d$ is a Gaussian vector with expectation vector m and covariance operator Σ.
- The characteristic function of random vector $X \in \mathbb{R}^d$ is

$$\boxed{\Phi_X(t) = e^{im^\top t - \frac{1}{2}t^\top \Sigma t},} \quad \forall t \in \mathbb{R}^d.$$

• $X = (X_1, \dots, X_d)^\top$ $\quad X_i \overset{iid}{\sim} \mathcal{N}(0,1)$

• $\Phi_X(t) =$

Proof as an exercise.

Theorem (Affine mapping stability)

If X is a Gaussian vector, then $Y = a + AX$ is also a Gaussian vector.

Proof as an exercise.

Gaussian Vectors

Theorem

The following assertions are equivalent:

- $X \in \mathbb{R}^d$ is a Gaussian vector with expectation vector m and covariance operator Σ.

- The characteristic function of random vector $X \in \mathbb{R}^d$ is

$$\Phi_X(t) = e^{im^\top t - \frac{1}{2} t^\top \Sigma t}, \quad \forall t \in \mathbb{R}^d.$$

$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

$= e^{i m_1 t_1 - \frac{1}{2} \sigma_1^2 t_1^2} \times e^{i m_2 t_2 - \frac{1}{2} \sigma_2^2 t_2^2}$

σ_{12}

$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$

$t^\top \Sigma t = \sigma_1^2 t_1^2 + \sigma_2^2 t_2^2$

Proof as an exercise.

Theorem (Affine mapping stability)

If X is a Gaussian vector, then $Y = a + AX$ is also a Gaussian vector.

Proof as an exercise.

- These results guarantee the existence of Gaussian vectors and provide a method to simulate them.

Gaussian Vectors

Proposition (Existence)

Let $Z = (Z_1, Z_2, \cdots, Z_d)^\top \in \mathbb{R}^d$ be a random vector with i.i.d. standard gaussian components. Then Z is a standard Gaussian vector.

Proof as an exercise

Gaussian Vectors

Proposition (Existence)

Let $Z = (Z_1, Z_2, \cdots, Z_d)^\top \in \mathbb{R}^d$ be a random vector with i.i.d. standard gaussian components. Then Z is a standard Gaussian vector.

Proof as an exercise

$d \times d$, rank $= k$

How can we simulate a Gaussian vector $X \sim N_d(m, \Sigma)$?

- $\Sigma = AA^T \quad A : d \times k$
- $Z_1, \ldots, Z_k \overset{iid}{\sim} N(0,1)$

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix} \sim N(0, I_k)$$

$$X = m + AZ \sim N\Big(m, \underbrace{A \operatorname{Cov}(Z) A^T}_{AA^T = \Sigma}\Big)$$

Gaussian Vectors

Proposition (Existence)

Let $Z = (Z_1, Z_2, \cdots, Z_d)^\top \in \mathbb{R}^d$ be a random vector with i.i.d. standard gaussian components. Then Z is a standard Gaussian vector.

Proof as an exercise

How can we simulate a Gaussian vector $X \sim N_d(m, \Sigma)$?

- Algorithm:
  1. Compute the Cholesky decomposition of $\Sigma = A\, A^\top$ with $A \in \mathbb{R}^{d \times k}$ is of rank k with $k \leq d$.
  2. Generate $Z_1, \ldots, Z_k$ are i.i.d. standard Gaussian rv's and set
  $$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_k \end{pmatrix} \in \mathbb{R}^k,$$
  3. Compute $m + AZ$.

Gaussian vectors

Theorem

Let $X = (X_1, \cdots, X_d)^\top \in \mathbb{R}^d$ be a Gaussian vector. X_i and X_j are independent if and only if $\mathrm{cov}(X_i, X_j) = 0$.

Proof is left as an exercise.

$X_1 \perp\!\!\!\perp X_2 \Rightarrow \mathrm{Cov}(X_1, X_2) = 0$

X Gaussian $\Leftarrow$

Gaussian vectors

Theorem

Let $X = (X_1, \cdots, X_d)^\top \in \mathbb{R}^d$ be a Gaussian vector. X_i and X_j are independent if and only if $\mathrm{cov}(X_i, X_j) = 0$.

Proof is left as an exercise.

- Independence property of Gaussian vectors is fundamental. It turns independence (a probabilistic notion) into a geometric one (orthogonality in $L_2(\mathbb{P})$).
- This is the core ingredient of the theory of Gaussian linear models.

Gaussian vectors

- In a nutshell, conditional Gaussian vectors are Gaussian vectors.

Theorem

Let $X = (X_1, \cdots, X_d)^\top \in \mathbb{R}^d$ be a Gaussian vector. Set $Y = (X_1, \cdots, X_m)^\top$ and $Z = (X_{m+1}, \cdots, X_d)^\top$. Then $Y|Z = z \sim N_d(m_c, \Sigma_c)$.
Assume that Σ_Z is positive definite. Then,

$$m_c = \mathbb{E}[Y|Z = z] = m_Y + \Sigma_{YZ}(z - m_Z)$$

and

$$\Sigma_c = \Sigma_Y - \Sigma_{YZ}\Sigma_Z^{-1}\Sigma_{YZ}$$

$m \times (d-m)$

$U = Y - AZ$

A s.t. $\mathrm{Cov}(U, Z) = 0$

$Y = U + AZ$

Question: ideas for the proof?

Cochran's Theorem

Definition (χ^2)

The chi-square distribution with ν degrees of freedom and noncentrality parameter μ (denoted by $\chi^2_{\nu,\mu}$) is the distribution of

$$\boxed{Z_1^2 + Z_2^2 + \cdots + Z_\nu^2,}$$

where $Z_1, \ldots, Z_\nu$ are i.i.d. Gaussian with unit variance and $\mathbb{E}Z_1 = \mu$, $\mathbb{E}Z_2 = \cdots = \mathbb{E}Z_\nu = 0$.

- If $\mu = 0$, we simply write $\chi^2_\nu = \chi^2_{\nu,0}$.

Cochran's Theorem

Definition (χ^2)

The chi-square distribution with ν degrees of freedom and noncentrality parameter μ (denoted by $\chi^2_{\nu,\mu}$) is the distribution of

$$Z_1^2 + Z_2^2 + \cdots + Z_\nu^2,$$

where $Z_1, \ldots, Z_\nu$ are i.i.d. Gaussian with unit variance and $\mathbb{E}Z_1 = \mu$, $\mathbb{E}Z_2 = \cdots = \mathbb{E}Z_\nu = 0$.
- If $\mu = 0$, we simply write $\chi^2_\nu = \chi^2_{\nu,0}$.

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \mathbb{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Theorem (simple version)

Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. Then

$$P_{\mathbb{1}_n}(X) = \bar{X}_n \mathbb{1}_n$$

1. $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i \sim N(\mu, \frac{\sigma^2}{n})$
2. $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ satisfies $\frac{n-1}{\sigma^2}\hat{\sigma}_n^2 \sim \chi^2(n-1)$
3. $\bar{X}_n \perp \hat{\sigma}_n^2$

$$\sigma^2 \chi^2(n-1)$$

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \| \overbrace{\underbrace{X - \bar{X}_n \mathbb{1}_n}_{P_{\mathbb{1}_n^\perp}(X)}}^{\sim N(0, \sigma^2 P_{\mathbb{1}_n^\perp})} \|^2$$

$$\mathrm{Cov}\left(P_{\mathbb{1}_n^\perp}(X), P_{\mathbb{1}_n}(X)\right)$$

Cochran's Theorem

$= P_{\mathbb{1}_n^\perp} \cdot \underbrace{\mathrm{Cov}(X,X)}_{= \sigma^2 I_n} P_{\mathbb{1}_n} = \sigma^2 P_{\mathbb{1}_n^\perp} P_{\mathbb{1}_n} = 0$

Theorem

Let $Y \in \mathbb{R}^n$ be a random vector with $\mathbb{E}[Y] = \mu$ and $\Sigma_Y = \sigma^2 I_n$ (isotropic) . Let $M_1, \ldots, M_k$ be mutually orthogonal linear subspaces of $\mathbb{R}^n$. Then

$\mathbb{R}^n = M_1 \oplus \cdots \oplus M_k$

i. $\boxed{P_{M_i}(Y)}$, *$i = 1, \ldots, k$, are uncorrelated,*

$$\mathbb{E}[P_{M_i}(Y)] = P_{M_i}(\mu), \quad \mathrm{Cov}(P_{M_i}(Y)) = \sigma^2 P_{M_i},$$

$\mathrm{Cov}(P_{M_i}(Y), P_{M_j}(Y)) = 0$

and

$$\mathbb{E}\|P_{M_i}(Y)\|^2 = \sigma^2 \dim(M_i) + \|P_{M_i}(\mu)\|^2.$$

Cochran's Theorem

Theorem

Let $Y \in \mathbb{R}^n$ be a random vector with $\mathbb{E}[Y] = \mu$ and $\Sigma_Y = \sigma^2 I_n$ (isotropic) . Let $M_1, \ldots, M_k$ be mutually orthogonal linear subspaces of $\mathbb{R}^n$. Then

i. *$P_{M_i}(Y)$, $i = 1, \ldots, k$, are uncorrelated,*

$$\mathbb{E}[P_{M_i}(Y)] = P_{M_i}(\mu), \quad \mathrm{Cov}(P_{M_i}(Y)) = \sigma^2 P_{M_i},$$

and

$$\mathbb{E}\|P_{M_i}(Y)\|^2 = \sigma^2 \mathrm{dim}(M_i) + \|P_{M_i}(\mu)\|^2.$$

If $Y \sim N_n(\mu, \sigma^2 I_n)$, then

ii. *$P_{M_i}(Y) \sim N_n(P_{M_i}(\mu), \sigma^2 P_{M_i})$ are mutually independent.*

iii. *$\|P_{M_i} Y\|^2$ are mutually independent and*

$$\|P_{M_i}(Y)\|^2 \sim \sigma^2 \chi^2_{\mathrm{dim}(M_i), \|P_{M_i}(\mu)\|/\sigma}, \quad 1 \leq i \leq k.$$