

# Answer TD1 GLM

Xiaopeng ZHANG

September 21, 2025

## Exercise 1 [Proof of Cochran's Theorem]

Let  $Z$  be a Gaussian random vector in  $\mathbb{R}^n$  with  $Z \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbb{I}_n)$ , where  $\boldsymbol{\mu} \in \mathbb{R}^n$  and  $\sigma > 0$ . Let  $F_1, \dots, F_m$  be subspaces of dimension  $d_i$ , orthogonal to each other such that  $\mathbb{R}^n = F_1 \oplus \dots \oplus F_m$ . For  $i = 1, \dots, m$ , let  $P_{F_i}$  denote the orthogonal projection matrix onto  $F_i$ . Prove that:

1. The random vectors  $P_{F_1}Z, \dots, P_{F_m}Z$  have respective distributions

$$\mathcal{N}(P_{F_1}\boldsymbol{\mu}, \sigma^2 P_{F_1}), \dots, \mathcal{N}(P_{F_m}\boldsymbol{\mu}, \sigma^2 P_{F_m})$$

2. The random vectors  $P_{F_1}Z, \dots, P_{F_m}Z$  are pairwise independent.

3. The random variables  $\frac{\|P_{F_1}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}, \dots, \frac{\|P_{F_m}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}$  have respective distributions  $\chi^2(d_1), \dots, \chi^2(d_m)$ .

4. The random variables  $\frac{\|P_{F_1}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}, \dots, \frac{\|P_{F_m}(Z-\boldsymbol{\mu})\|^2}{\sigma^2}$  are pairwise independent.

1. By the linearity of expectation and the properties of Gaussian distributions, we have:

$$\begin{aligned}\mathbb{E}[P_{F_i}Z] &= P_{F_i}\mathbb{E}[Z] = P_{F_i}\boldsymbol{\mu}, \\ \text{Cov}(P_{F_i}Z) &= P_{F_i}\text{Cov}(Z)P_{F_i}^T = \sigma^2 P_{F_i}.\end{aligned}$$

Now we need to show that  $P_{F_i}Z$  also has a Gaussian distribution. We can't use the result that state linear transformations of Gaussian vectors are Gaussian, because we are proving this result. To do this, we can use the characteristic function of the Gaussian distribution. The characteristic function of a Gaussian random vector  $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$  is given by:

$$\phi_X(t) = \exp\left(it^T \boldsymbol{\mu} - \frac{1}{2}t^T \Sigma t\right).$$

For the random vector  $P_{F_i}Z$ , we have:

$$\begin{aligned}
\phi_{P_{F_i}Z}(t) &= \mathbb{E} \left[ e^{it^T P_{F_i}Z} \right] \\
&= \mathbb{E} \left[ e^{it^T P_{F_i}(\mu + \sigma W)} \right] \quad (\text{where } W \sim \mathcal{N}(0, \mathbb{I}_n)) \\
&= e^{it^T P_{F_i}\mu} \mathbb{E} \left[ e^{it^T P_{F_i}\sigma W} \right] \\
&= e^{it^T P_{F_i}\mu} \mathbb{E} \left[ e^{i(P_{F_i}^T t)\sigma W} \right] \\
&= e^{it^T P_{F_i}\mu} \cdot \exp \left( -\frac{1}{2} \|P_{F_i}^T t\|^2 \sigma^2 \right) \\
&= \exp \left( it^T P_{F_i}\mu - \frac{1}{2} \|P_{F_i}^T t\|^2 \sigma^2 \right).
\end{aligned}$$

This shows that  $P_{F_i}Z$  has a Gaussian distribution with mean  $P_{F_i}\mu$  and covariance  $\sigma^2 P_{F_i}$ .

2. To show that the random vectors  $P_{F_i}Z$  and  $P_{F_j}Z$  are independent for  $i \neq j$ , we can start by calculating the covariance, then we can prove that they are jointly Gaussian. Two jointly Gaussian random vectors are independent if and only if their covariance is zero. We have:

$$\begin{aligned}
\text{Cov}(P_{F_i}Z, P_{F_j}Z) &= \mathbb{E}[(P_{F_i}Z - \mathbb{E}[P_{F_i}Z])(P_{F_j}Z - \mathbb{E}[P_{F_j}Z])^T] \\
&= \mathbb{E}[(P_{F_i}(Z - \mu))(P_{F_j}(Z - \mu))^T] \\
&= P_{F_i} \mathbb{E}[(Z - \mu)(Z - \mu)^T] P_{F_j}^T \\
&= P_{F_i}(\sigma^2 \mathbb{I}_n) P_{F_j}^T \\
&= \sigma^2 P_{F_i} P_{F_j}^T.
\end{aligned}$$

Since  $F_i$  and  $F_j$  are orthogonal subspaces, we have  $P_{F_i}P_{F_j} = 0$ .

Therefore,  $\text{Cov}(P_{F_i}Z, P_{F_j}Z) = 0$ , now it is left to be shown that  $P_{F_i}Z$  and  $P_{F_j}Z$  are jointly Gaussian.

To show that  $P_{F_i}Z$  and  $P_{F_j}Z$  are jointly Gaussian, we need to use their characteristic functions. The characteristic function of a Gaussian random vector is given by:

$$\begin{aligned}
\phi_{P_{F_i}Z}(t) &= \exp \left( it^T P_{F_i}\mu - \frac{1}{2} \|P_{F_i}^T t\|^2 \sigma^2 \right), \\
\phi_{P_{F_j}Z}(t) &= \exp \left( it^T P_{F_j}\mu - \frac{1}{2} \|P_{F_j}^T t\|^2 \sigma^2 \right).
\end{aligned}$$

Their joint characteristic function is given by:

$$\begin{aligned}
\phi_{P_{F_i}Z, P_{F_j}Z}(t_1, t_2) &= \mathbb{E} \left[ e^{it_1^T P_{F_i}Z + it_2^T P_{F_j}Z} \right] \\
&= \mathbb{E} \left[ e^{it_1^T P_{F_i}(\mu + \sigma W) + it_2^T P_{F_j}(\mu + \sigma W)} \right] \\
&= e^{it_1^T P_{F_i}\mu + it_2^T P_{F_j}\mu} \mathbb{E} \left[ e^{i(t_1^T P_{F_i} + t_2^T P_{F_j})\sigma W} \right] \\
&= e^{it_1^T P_{F_i}\mu + it_2^T P_{F_j}\mu} \cdot \exp \left( -\frac{1}{2} \|(t_1^T P_{F_i} + t_2^T P_{F_j})\sigma W\|^2 \right) \\
&= \exp \left( it_1^T P_{F_i}\mu + it_2^T P_{F_j}\mu - \frac{1}{2} \|(t_1^T P_{F_i} + t_2^T P_{F_j})\sigma W\|^2 \right).
\end{aligned}$$

We thus have shown that  $P_{F_i}Z$  and  $P_{F_j}Z$  are jointly Gaussian. Since their covariance is zero, they are independent.

3. We know that  $P_{F_i}Z \sim \mathcal{N}(P_{F_i}\boldsymbol{\mu}, \sigma^2 P_{F_i})$ . Let  $Y_i = P_{F_i}Z - P_{F_i}\boldsymbol{\mu}$ . Then,  $Y_i \sim \mathcal{N}(0, \sigma^2 P_{F_i})$ . The matrix  $P_{F_i}$  is a projection matrix onto a subspace of dimension  $d_i$ , so it has rank  $d_i$ . Therefore, we can write  $P_{F_i} = U_i U_i^T$ , where  $U_i$  is an  $n \times d_i$  matrix whose columns form an orthonormal basis for the subspace  $F_i$ . The  $\chi^2$  distribution with  $k$  degrees of freedom can be defined as the distribution of the sum of the squares of  $k$  independent standard normal random variables. To show that  $\frac{\|Y_i\|^2}{\sigma^2} \sim \chi^2(d_i)$ , we can express  $Y_i$  in terms of a standard normal vector. Let  $W \sim \mathcal{N}(0, \mathbb{I}_n)$  be a standard normal vector in  $\mathbb{R}^n$ . Then, we can write:

$$\begin{aligned} Y_i &= P_{F_i}Z - P_{F_i}\boldsymbol{\mu} \\ &= P_{F_i}(\boldsymbol{\mu} + \sigma W) - P_{F_i}\boldsymbol{\mu} \\ &= P_{F_i}\sigma W \\ &= \sigma P_{F_i}W. \end{aligned}$$

Therefore, we have:

$$\begin{aligned} \frac{\|Y_i\|^2}{\sigma^2} &= \frac{\sigma^2 \|P_{F_i}W\|^2}{\sigma^2} \\ &= W^T P_{F_i}^T P_{F_i} W \\ &= W^T P_{F_i} W \\ &\sim \chi^2(d_i). \end{aligned}$$

Because  $P_{F_i}$  is an orthogonal projection matrix onto a subspace of dimension  $d_i$ ,  $P_{F_i}^T = P_{F_i}$ .

4. Since we have already shown that the random vectors  $P_{F_i}Z$  and  $P_{F_j}Z$  are independent for  $i \neq j$ , it follows that any functions of these independent random vectors are also independent. In particular, the random variables  $\frac{\|P_{F_i}(Z - \boldsymbol{\mu})\|^2}{\sigma^2}$  and  $\frac{\|P_{F_j}(Z - \boldsymbol{\mu})\|^2}{\sigma^2}$  are functions of the independent random vectors  $P_{F_i}Z$  and  $P_{F_j}Z$ , respectively. Therefore, these random variables are also independent for  $i \neq j$ .
5. **Final note:** characteristic function of any distribution is:

$$\phi_Z(t) = \mathbb{E}(e^{i\langle t, Z \rangle})$$

### Properties of Fourier Transform:

**General Formalization:** For a function  $f : \mathbb{R} \rightarrow \mathbb{C}$ , the Fourier transform is defined as:

$$\hat{f}(\xi) = \mathcal{F}[f](\xi) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i \xi t} dt$$

The inverse Fourier transform is:

$$f(t) = \mathcal{F}^{-1}[\hat{f}](t) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi i \xi t} d\xi$$

### Key Properties:

- (a) **Linearity:**  $\mathcal{F}[af + bg] = a\mathcal{F}[f] + b\mathcal{F}[g]$
- (b) **Time shifting:**  $\mathcal{F}[f(t - a)](\xi) = e^{-2\pi i a \xi} \hat{f}(\xi)$
- (c) **Frequency shifting:**  $\mathcal{F}[e^{2\pi i a t} f(t)](\xi) = \hat{f}(\xi - a)$
- (d) **Scaling:**  $\mathcal{F}[f(at)](\xi) = \frac{1}{|a|} \hat{f}\left(\frac{\xi}{a}\right)$
- (e) **Conjugation:**  $\mathcal{F}[\overline{f(t)}](\xi) = \overline{\hat{f}(-\xi)}$
- (f) **Time reversal:**  $\mathcal{F}[f(-t)](\xi) = \hat{f}(-\xi)$
- (g) **Differentiation:**  $\mathcal{F}[f'(t)](\xi) = 2\pi i \xi \hat{f}(\xi)$
- (h) **Integration:**  $\mathcal{F}\left[\int_{-\infty}^t f(\tau) d\tau\right](\xi) = \frac{\hat{f}(\xi)}{2\pi i \xi}$
- (i) **Convolution theorem:**  $\mathcal{F}[(f * g)(t)](\xi) = \hat{f}(\xi) \hat{g}(\xi)$
- (j) **Parseval's theorem:**  $\int_{-\infty}^{\infty} |f(t)|^2 dt = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 d\xi$
- (k) **Plancherel's theorem:**  $\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle$

**Relation to Fourier transform (answer to the question).** The characteristic function is the Fourier transform of the law (probability measure) of  $Z$ . If  $Z$  has a density  $f_Z$  on  $\mathbb{R}^n$ :

$$\phi_Z(t) = \int_{\mathbb{R}^n} e^{i t^\top x} f_Z(x) dx.$$

Using the “angular-frequency” convention  $\mathcal{F}_\omega[f](\omega) = \int f(x) e^{-i \omega^\top x} dx$ , one has

$$\phi_Z(t) = \mathcal{F}_\omega[f_Z](-t).$$

With the  $2\pi$ -normalized convention,

$$\hat{f}_Z(\xi) = \int f_Z(x) e^{-2\pi i \xi^\top x} dx, \quad \Rightarrow \quad \phi_Z(t) = \hat{f}_Z\left(-\frac{t}{2\pi}\right).$$

When a density exists, an inversion formula is

$$f_Z(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i t^\top x} \phi_Z(t) dt,$$

with the constant adjusted to the chosen Fourier convention.

## Exercise 2 Proof of Proposition 1 Chapter 1

Let  $X$  be the design matrix of size  $n \times (p+1)$ . Assume  $X$  is full rank, i.e.,  $\text{rank}(X) = p+1$ . Consider the linear model

$$Y = X\beta + \varepsilon,$$

with  $\beta \in \mathbb{R}^{p+1}$ . Define the ordinary least squares estimator (OLSE)

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|^2.$$

1. Show that

$$\hat{\beta} = \hat{\beta}(Y) = (X^\top X)^{-1} X^\top Y.$$

2. Application for  $p = 1$ : let  $(x_1, y_1), \dots, (x_n, y_n)$  be  $n$  pairs of real numbers. Determine the real  $\hat{a}$  and  $\hat{b}$  that minimize

$$\text{RSS}(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Interpret the result.

1. We want to minimize the function  $\beta \mapsto \|Y - X\beta\|^2$ .

$$\|Y - X\beta\| \leq \|Y\| + \|X\beta\| \leq \|Y\| + C\|\beta\|$$

As this shows that the function goes to infinity as  $\|\beta\|$  goes to infinity, the minimum is attained at some point  $\hat{\beta}$ . The function is differentiable and convex, so the minimum is attained at a point where the gradient is zero.

calculating the gradient, we have

$$\nabla_{\beta} \|Y - X\beta\|^2 = -2X^\top (Y - X\beta)$$

Setting this to zero, we have

$$X^\top Y = X^\top X \hat{\beta}$$

As  $X$  is full rank,  $X^\top X$  is invertible, and we have

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y$$