Answer TD1 GLM

Xiaopeng ZHANG

September 30, 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 $F_1, \ldots, F_m$ be subspaces of dimension d_i, orthogonal to each other such that $\mathbb{R}^n = F_1 \oplus \cdots \oplus F_m$. For $i = 1, \ldots, m$, let P_{F_i} denote the orthogonal projection matrix onto F_i. Prove that:

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

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

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

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

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

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

$$\mathbb{E}[P_{F_i}Z] = P_{F_i}\mathbb{E}[Z] = P_{F_i}\boldsymbol{\mu},$$
$$\mathrm{Cov}(P_{F_i}Z) = P_{F_i}\mathrm{Cov}(Z)P_{F_i}^T = \sigma^2 P_{F_i}.$$

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:

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

This shows that $P_{F_i}Z$ has a Gaussian distribution with mean $P_{F_i}\boldsymbol{\mu}$ 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:

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

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

Therefore, $\mathrm{Cov}(P_{F_i}Z, P_{F_j}Z) = 0$, 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:

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

Their joint characteristic function is given by:

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

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 χ^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 $f : \mathbb{R} \to \mathbb{C}$, 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 2π-normalized convention,

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

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 1.1 (Proof of Cochran's theorem)**.** Let Z be a Gaussian random vector in $\mathbb{R}^n$ with $Z \sim N(\mu, \sigma^2 I_n)$, where $\mu \in \mathbb{R}^n$ and $\sigma > 0$. Let $F_1, \ldots, F_m$ be subspaces of dimension d_i, orthogonal to each other such that $\mathbb{R}^n = F_1 \oplus \cdots \oplus F_m$. For $i = 1, \ldots, m$, let P_{F_i} denote the orthogonal projection matrix onto F_i. Prove that

1. The random vectors $P_{F_1}Z, \ldots, P_{F_m}Z$ have respective distributions
$$N(P_{F_1}\mu, \sigma^2 P_{F_1}), \ldots, N(P_{F_m}\mu, \sigma^2 P_{F_m}) \tag{1}$$

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

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

4. The random variables

$$\frac{\|P_{F_1}(Z-\mu)\|^2}{\sigma^2},\ldots,\frac{\|P_{F_m}(Z-\mu)\|^2}{\sigma^2} \tag{3}$$

are pairwise independent.

Exercise 1.2 (Proof of Proposition 1. of the chapter 1)**.** Let X be the design matrix of size $n\times(p+1)$. We assume X to be full rank ($\text{rank}(X)=p+1$). Let define the following linear model

$$Y = X\beta + \epsilon$$

with $\beta\in\mathbb{R}^{p+1}$. Let

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

be the ordinary least square estimator (OLSE).

1. Show that OLSE exists and is unique such that

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

2. Application for $p=1$: Let $(x_1,y_1),\ldots,(x_n,y_n)$ 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.

Exercise 1.3. Let X be a $n\times p$ matrix of rank p. Let $\hat{Y}$ be the orthogonal project on the space $[X]$ generated by the column vectors of X of a vector Y of $\mathbb{R}^n$. Show that $\sum_{i=1}^n(Y_i-\hat{Y}_i)=0$ if one of the column vectors of X is the vector $\mathbf{1}_n=(1,\ldots,1)$. Interpret.

Exercise 1.4. We consider the following simple linear regression statistical model: $Y_i = \beta x_i + \varepsilon_i$, for $i=1,\ldots,n$ where the ε_i are independent, centered, of constant variance. We define two estimators of $\beta\in\mathbb{R}$:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_iY_i}{\sum_{i=1}^n x_i^2} \quad \text{and} \quad \beta^\star = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}$$

1. What is the logic of construction of these estimators?
2. Show that they are unbiased estimators of β.
3. Compare the variances of these two estimators.

Exercise 1.5 (An important result)**.** We consider the Gaussian linear regression model:

$$Y = X\beta + \epsilon, \quad \epsilon\sim N(0_n,\sigma^2 I_n)$$

where $\beta\in\mathbb{R}^r$, $Y\in\mathbb{R}^n$ and X matrix of size $n\times r$ of rank r.

1. Recall the matrix closed form of the OLSE and give an unbiased estimator of $\sigma^2>0$.
2. Compute the maximum likelihood estimators of β and σ^2.
3. Conclude.

Exercise 1.6 (Unbiased estimator of σ^2 in the non-Gaussian model)**.** Consider the following non-Gaussian linear model:

$$Y = X\beta + \epsilon$$

with $\beta \in \mathbb{R}^p$, X of full rank, and the ϵ_i independent, centered and of variance σ^2. We pose:

$$\hat{\sigma}^2 = \frac{1}{n-p}\|Y - X\hat{\beta}\|^2$$

We note $\mathrm{Tr}(\cdot)$ the trace of a matrix.

1. Show that $(n-p)\hat{\sigma}^2 = \mathrm{Tr}(\epsilon^\top P_{X^\perp}\epsilon)$
2. Using the fact that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for A and B of respective size $(m \times n)$ and $(n \times m)$, show that
$$(n-p)E_\beta[\hat{\sigma}^2] = \sigma^2\mathrm{Tr}(P_{X^\perp})$$
3. Deduce that $E_\beta[\hat{\sigma}^2] = \sigma^2$.

Exercise 1.7 (Proof of theorem 4 chapter 4)**.** Consider the following Gaussian linear model $Y = X\beta + \epsilon$ where $\beta \in \mathbb{R}^r$, X is a full rank matrix of size $n \times r$ $(n > r)$. Let $C \in M_{q,r}(\mathbb{R})$. We want to test

$$H_0 : C\beta = 0_q \quad \text{versus} \quad H_1 : C\beta \neq 0_q$$

We assume that $\mathrm{rg}(C) = q \leq r$. Therefore, you will note that $\mathrm{rg}(C^\top) = q$ where $C^\top$ is the transpose of C.

1. Show that if $Z \sim N_q(0_q, \Sigma)$ then $Z^\top\Sigma^{-1}Z \sim \chi^2_q$.
2. Show that $C(X^\top X)^{-1}C^\top$ is a symmetric and invertible matrix.
3. Recall the ordinary least squares expression $\hat{\beta}$.
4. What is the law of $\hat{\beta}$?
5. Deduce the law of $C\hat{\beta}$ under the hypothesis H_0.
6. Deduce that, under H_0,
$$R = \frac{(C\hat{\beta})^\top(C(X^\top X)^{-1}C^\top)^{-1}(C\hat{\beta})}{\sigma^2} \sim \chi^2_q$$
7. Conclude that, under H_0,
$$F = \frac{\hat{\beta}^\top C^\top(C(X^\top X)^{-1}C^\top)^{-1}C\hat{\beta}}{q\hat{\sigma}^2}$$
is distributed according to a Fisher distribution with $(q, n-r)$ degrees of freedom. Each step of the reasoning must be carefully justified.
8. Justify and construct a test of H_0 against H_1 of level α.

Exercise 1.8 (MCQ)**.** We have observations $(x_i, y_i) \in \mathbb{R}^2$, $\forall i = 1, \dots, n$. We consider the following classical Gaussian linear model:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \forall i = 1, \dots, n$$

where $(\beta_0, \beta_1) \in \mathbb{R}^2$ and $\varepsilon_i \sim N(0, \sigma^2)$ are i.i.d.

Let $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$. Assume X is a full rank matrix and note $\hat{\beta}_0$ and $\hat{\beta}_1$ the least squares estimators of β_0 and β_1.

For each of the following questions, give the answer.

1. Are the variables Y_i independent and identically distributed?

   a) Yes b) No c) not always

2. Does the regression line calculated on the observations pass through the mean point $(\bar{x}, \bar{y})$?

   a) Yes b) No c) Only if I am lucky

3. Is it possible to find estimators of β_0 and β_1 with smaller variance than the ordinary least squares estimators?

   a) Yes b) No c) Maybe.

4. Are $\hat{\beta}_0$ and $\hat{\beta}_1$ independent?

   a) Yes b) No c) It depends on the matrix X

5. If the coefficient of determination R^2 calculated on the observations is equal to 1, are the points $(x_i, y_i)_{i=1,\dots,n}$ aligned?

   a) Yes b) No c) Not necessarily

6. Are $\hat{Y}$ and $Y - \hat{Y}$ independent?

   a) Yes b) No c) It depends on the matrix X

7. Are $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$ and $Y - \hat{Y}$ independent?

   a) Yes b) No c) It depends on the matrix X

8. Is the maximum likelihood estimator of σ^2 unbiased?

   a) Yes b) No c) We don't know

Exercise 1.9 (This exercise will be solved without the tools of linear algebra)**.** Let $(x_1, y_1), \dots, (x_n, y_n)$ be n pairs of real numbers. We suppose that y_i are the realization of Y_i whose law is given by the following equation:

$$Y_i = a + bx_i + \varepsilon_i, \quad \varepsilon_i \sim_{i.i.d.} N(0, \sigma^2)$$

1. Determine $\hat{A}$ and $\hat{B}$ the maximum likelihood estimators of a and b. Interpret the estimators.

2. Show that these estimators are unbiased.

3. Calculate the variance of the estimators $\mathrm{Var}_\beta(\hat{A})$ and $\mathrm{Var}_\beta(\hat{B})$. How do these variances vary as a function of σ^2 and the experimental design $x_1, \ldots, x_n$?

4. Compute the covariance of $\hat{A}$ and $\hat{B}$. Comment.

5. Let $\hat{Y}_i = \hat{A} + \hat{B}x_i$ and $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$. Show that $\sum_{i=1}^n \hat{\varepsilon}_i = 0$.

6. Show that $\frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-2}$ is an unbiased estimator of σ^2.

7. Let x_{n+1} be another value. We define $\hat{Y}_{n+1} = \hat{A} + \hat{B}x_{n+1}$. Compute the variance of this prediction.

8. Furthermore, let $Y_{n+1} = A + Bx_{n+1} + \varepsilon_{n+1}$. Calculate the variance of $\hat{\varepsilon}_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$. Compare it to the variance of ε_i (for $i = 1, \ldots, n$).

9. Gauss-Markov Theorem:

   (a) Show that $\hat{B}$ is written as a linear combination of the observations (we will explain the weights).

   (b) Consider $\tilde{B} = \sum_{i=1}^n \lambda_i Y_i$ another unbiased estimator of B, written as a linear combination of Y_i. Show that $\sum_{i=1}^n \lambda_i = 0$ and $\sum_{i=1}^n \lambda_i x_i = 1$.

   (c) Deduce that $\mathrm{Var}_\beta(\tilde{B}) \geq \mathrm{Var}_\beta(\hat{B})$