Annette J. Dobson

An Introduction to Generalized Linear Models

Solutions

by Konstantinos Kirillov

CHAPTER 1

Exercise 1.1:

Let $\boldsymbol{Y_1}$ and $\boldsymbol{Y_2}$ be independent random variables with $\boldsymbol{Y_1 \sim N(1,3)}$ and $\boldsymbol{Y_2 \sim N(2,5)}$. If $\boldsymbol{W_1 = Y_1 + 2Y_2}$ and $\boldsymbol{W_2 = 4Y_1 - Y_2}$ what is the joint distribution of $\boldsymbol{W_1}$ and $\boldsymbol{W_2}$?

SOLUTION:

A reminder from the book:

1.4.1 Normal distributions:

1. If the random variable Y has the Normal distribution with mean μ and variance σ^2, its probability density function is:

$$f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma^2}\right)^2\right]$$

   We denote this by $Y \sim N(\mu,\sigma^2)$.

2. The Normal distribution with $\mu = 0$ and $\sigma^2 = 1$, $Y \sim N(0,1)$, is called the **standard Normal distribution**.

3. Let $Y_1, \dots, Y_n$ denote Normally distributed random variables with $Y_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1,\dots,n$ and let the covariance of Y_i and Y_j be denoted by:

$$cov(Y_i, Y_j) = \rho_{ij}\sigma_i\sigma_j,$$

   where ρ_{ij} is the correlation coefficient for Y_i and Y_j. Then the joint distribution of the Y_i's is the **multivariate Normal distribution** with mean vector $\mu = [\mu_1, \dots, \mu_n]^T$ and variance-covariance matrix V with diagonal elements σ_i^2 and non-diagonal elements $\rho_{ij}\sigma_i\sigma_j$ for $i \neq j$. We write this as:

$$\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V}), \text{ where } \boldsymbol{y} = [Y_1, \dots, Y_n]^T$$

4. Suppose the random variables $Y_1, \dots, Y_n$ are independent and normally distributed with the distributions $Y_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1,\dots,n$. If

$$W = a_1 Y_1 + a_2 Y_2 + \cdots + a_n Y_n,$$

where the a_i's are constants. Then W is also Normally distributed, so that:

$$W = \sum_{i=1}^{n} a_i Y_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

It seems that the joint distribution of two normally distributed variables is yet another normal distribution. In this exercise, in order to find the joint distribution of W_1 and W_2, we first need to determine the mean, the variance and the covariance of W_1 and W_2 and then use those to derive the joint distribution.

<u>Given that:</u>
$Y_1 \sim N(1, 3)$
$Y_2 \sim N(2, 5)$

First, let us find the means of W_1 and W_2:

$E(W_1) = E(Y_1 + 2Y_2) = E(Y_1) + 2E(Y_2) = 1 + 2 \cdot 2 = 5 \Rightarrow \boldsymbol{E(W_1) = 5}$

$E(W_2) = E(4Y_1 - Y_2) = 4E(Y_1) - E(Y_2) = 4 \cdot 1 - 2 = 2 \Rightarrow \boldsymbol{E(W_2) = 2}$

Next, let us calculate the variances of W_1 and W_2:

$Var(W_1) = Var(Y_1 + 2Y_2) = Var(Y_1) + 2^2 \cdot Var(Y_2) = 3 + 4 \cdot 5 = 23 \Rightarrow \boldsymbol{Var(W_1) = 23}$

$Var(W_2) = Var(4Y_1 - Y_2) = 4^2 \cdot Var(Y_1) + Var(Y_2) = 16 \cdot 3 + 5 = 53 \Rightarrow \boldsymbol{Var(W_2) = 53}$

And finally, let us also compute the covariance between W_1 and W_2:

$Cov(W_1, W_2) = Cov(Y_1 + 2Y_2, 4Y_1 - Y_2) =$
$= Cov(Y_1,\ 4Y_1) + Cov(Y_1, -Y_2) + Cov(2Y_2,\ 4Y_1) + Cov(2Y_2, -Y_2) =$
$= 4Var(Y_1) - Cov(Y_1, Y_2) + 8Cov(Y_2,\ Y_1) - 2Var(Y_2) =$
$= 4 \cdot 3 - 0 + 8 \cdot 0 - 2 \cdot 5 = 2 \Rightarrow$
$\Rightarrow \boldsymbol{Cov(W_1, W_2) = 2}$

Therefore, the joint distribution will be:

$$\begin{pmatrix} \boldsymbol{W_1} \\ \boldsymbol{W_2} \end{pmatrix} \sim \boldsymbol{N}\left[\begin{pmatrix} \mathbf{5} \\ \mathbf{2} \end{pmatrix}, \begin{pmatrix} \mathbf{23} & \mathbf{2} \\ \mathbf{2} & \mathbf{53} \end{pmatrix}\right]$$

The correlation coefficient between W_1 and W_2 in this case shall be:

$$\rho = \frac{Cov(W_1, W_2)}{\sigma_{W_1} \cdot \sigma_{W_2}} = \frac{Cov(W_1, W_2)}{\sqrt{Var(W_1)} \cdot \sqrt{Var(W_2)}} = \frac{2}{\sqrt{23} \cdot \sqrt{53}} \approx \frac{2}{4.8 \cdot 7.3} \approx 0.057 \Rightarrow \boldsymbol{\rho = 0.057}$$

Therefore, another way to express the joint distribution, would be:

$$f(W_1, W_2) = \frac{1}{2 \cdot \pi \cdot \sigma_{W_1} \cdot \sigma_{W_2} \cdot \sqrt{1 - \rho^2}} exp\left[-\frac{Z_{W_1}^2 - 2 \cdot Z_{W_1} \cdot Z_{W_2} + Z_{W_2}^2}{2 \cdot \sqrt{1 - \rho^2}}\right] \Rightarrow$$

$$\Rightarrow \boldsymbol{f(W_1, W_2)} = \frac{\mathbf{1}}{\mathbf{2} \cdot \boldsymbol{\pi} \cdot \mathbf{4.8} \cdot \mathbf{7.3} \cdot \sqrt{\mathbf{1} - \mathbf{0.057}^2}} \boldsymbol{exp}\left[-\frac{\boldsymbol{Z_{W_1}^2} - \mathbf{2} \cdot \boldsymbol{Z_{W_1}} \cdot \boldsymbol{Z_{W_2}} + \boldsymbol{Z_{W_2}^2}}{\mathbf{2} \cdot \sqrt{\mathbf{1} - \mathbf{0.057}^2}}\right]$$

Where:

$$Z_{W_1} = \frac{W_1 - \mu_{W_1}}{\sigma_{W_1}}$$

$$Z_{W_2} = \frac{W_2 - \mu_{W_2}}{\sigma_{W_2}}$$

Exercise 1.2:

Let Y_1 and Y_2 be independent random variables with $Y_1 \sim N(0,1)$ and $Y_2 \sim N(3,4)$.
a. What is the distribution of Y_1^2?
b. If $y = \begin{bmatrix} Y_1 \\ (Y_2 - 3)/2 \end{bmatrix}$, obtain an expression for $y^T y$. What is its distribution?
c. If $y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ and its distribution is $y \sim N(\mu, V)$, obtain an expression for $y^T V^{-1} y$. What is its distribution?

SOLUTION:

A reminder from the book:

1.4.2 Chi-squared distribution:

1. The **central chi-squared distribution** with n degrees of freedom is defined as the sum of squares of n independent random variables $Z_1, \ldots, Z_n$ each with the standard Normal distribution. It is denoted by:

$$X^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

   In matrix notation, if $\mathbf{z} = [Z_1, \ldots, Z_n]^T$ then $\mathbf{z}^T\mathbf{z} = \sum_{i=1}^{n} Z_i^2$ so that $X^2 = \mathbf{z}^T\mathbf{z} \sim \chi_n^2$.

2. If X^2 has the distribution χ_n^2, then its expected value is $E(X^2) = n$ and its variance is $Var(X^2) = 2n$.

3. If $Y_1, \ldots, Y_n$ are independent Normally distributed random variables each with the distribution $Y_i \sim N(\mu_i, \sigma_i^2)$ then:

$$X^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_n^2$$

   because each of the variables $Z_i = (Y_i - \mu_i)/\sigma_i$ has the standard Normal distribution $N(0,1)$.

4. Let $Z_1, \ldots, Z_n$ be independent random variables each with the distribution $N(0,1)$ and let $Y_i = Z_i + \mu_i$, where at least one of the μ_i's is non-zero. Then the distribution of:

$$\sum Y_i^2 = \sum (Z_i + \mu_i)^2 = \sum Z_i^2 + 2\sum Z_i\mu_i + \sum \mu_i^2$$

has larger mean $n + \lambda$ and larger variance $2n + 4\lambda$ than χ_n^2 where $\lambda = \sum \mu_i^2$. This is called the **non-central chi-squared distribution** with n degrees of freedom and **non-centrality parameter** λ. It is denoted by $\chi_n^2(\lambda)$.

5. Suppose that the Y_i's are not necessarily independent and the vector $\boldsymbol{y} = [Y_1, \dots, Y_n]^T$ has the multivariate normal distribution $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$ where the variance-covariance matrix $\boldsymbol{V}$ is non-singular and its inverse is $\boldsymbol{V}^{-1}$. Then:

$$X^2 = (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \sim \chi_n^2$$

6. More generally if $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$ then the random variable $\boldsymbol{y}^T \boldsymbol{V}^{-1} \boldsymbol{y}$ has the non-central chi-squared distribution $\chi_n^2(\lambda)$ where $\lambda = \boldsymbol{\mu}^T \boldsymbol{V}^{-1} \boldsymbol{\mu}$.

7. If $X_1^2, \dots, X_m^2$ are m independent random variables with the chi-squared distributions $X_i^2 \sim \chi_{n_i}^2(\lambda_i)$, which may or may not be central, then their sum also has a chi-squared distribution with $\sum n_i$ degrees of freedom and non-centrality parameter $\sum \lambda_i$, i.e.,

$$\sum_{i=1}^{m} X_i^2 \sim \chi^2_{\sum_{i=1}^{m} n_i}\left(\sum_{i=1}^{m} \lambda_i\right)$$

This is called the **reproductive property** of the chi-squared distribution.

8. Let $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{y}$ has n elements but the Y_i's are not independent so that $\boldsymbol{V}$ is singular with rank $k < n$ and the inverse of $\boldsymbol{V}$ is not uniquely defined. Let $\boldsymbol{V}^-$ denote a generalized inverse of $\boldsymbol{V}$. Then the random variable $\boldsymbol{y}^T \boldsymbol{V}^- \boldsymbol{y}$ has the non-central chi-squared distribution with k degrees of freedom and non-centrality parameter $\lambda = \boldsymbol{\mu}^T \boldsymbol{V}^- \boldsymbol{\mu}$.

a. As property 1 from above would suggest, the chi-squared distribution with n degrees of freedom, χ_n^2, is the distribution of the sum of the squares of n independent standard normal random variables. If Y_1 is a random variable following a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ ($Y_1 \sim N(0,1)$), then the distribution of Y_1^2 is a special case of the **chi-squared distribution with one degree of freedom, $\boldsymbol{\chi_1^2}$**. Meaning that:

$$\boldsymbol{Y_1^2 \sim \chi_1^2}$$

The chi-squared distribution with 1 degree of freedom is sometimes referred to as the exponential distribution with rate parameter $\lambda = 2$ ($mean = 1/\lambda = 1/2$, $variance = 1/\lambda^2 = 1/4$).

So, the distribution of Y_1^2 is χ_1^2 or equivalently, an exponential distribution with rate parameter $\lambda = 2$.

b. The expression $y^T y$ is the dot product of the vector y with itself. So:

$$y^T y = \begin{bmatrix} Y_1 & \frac{Y_2 - 3}{2} \end{bmatrix} \begin{bmatrix} Y_1 \\ \frac{Y_2 - 3}{2} \end{bmatrix} = Y_1^2 + \left(\frac{Y_2 - 3}{2}\right)^2 \Rightarrow \boldsymbol{y^T y = Y_1^2 + \frac{Y_2^2 - 6 \cdot Y_2 + 9}{4}}$$

We know that $Y_1 \sim N(0,1)$ and $Y_2 \sim N(3,4)$, and that they are independent. We also know (form a) that Y_1^2 is a special case of the chi-squared distribution with one degree of freedom, χ_1^2, or in other words: $Y_1^2 \sim \chi_1^2$.

Furthermore, we are given that: $Y_2 \sim N(3,4)$, thus:

$$Y_2 \sim N(3,4) \Rightarrow Y_2 - 3 \sim N(0,4) \Rightarrow \frac{Y_2 - 3}{2} \sim N(0,1) \Rightarrow \left(\frac{Y_2 - 3}{2}\right)^2 \sim \chi_1^2$$

Since both Y_1^2 and $\left(\frac{Y_2-3}{2}\right)^2$ are independent and follow a chi-squared distribution with 1 degree of freedom, then it follows that their sum will also follow the chi-squared distribution, but with two degrees of freedom, that are coming from the two terms combined. Therefore (and also according to property 3):

$$\boldsymbol{y^T y = Y_1^2 + \frac{Y_2^2 - 6 \cdot Y_2 + 9}{4} \sim \chi_2^2}$$

c. We know that $Y_1 \sim N(0,1)$ and $Y_2 \sim N(3,4)$. Given that: $y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ and its distribution is $y \sim N(\mu, V)$, we have that:

The mean vector $\boldsymbol{\mu}$ of y, is:

$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

While the Variance-Covariance matrix V, is a diagonal matrix, because Y_1 and Y_2 are independent and it is:

$\boldsymbol{V} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$

Let us also compute the inverse of Variance-Covariance matrix $\boldsymbol{V}$, $\boldsymbol{V^{-1}}$ as it will be used:

$$\boldsymbol{V^{-1}} = \frac{1}{1\cdot 4 - 0\cdot 0}\begin{bmatrix}4 & 0\\ 0 & 1\end{bmatrix} = \frac{1}{4}\begin{bmatrix}4 & 0\\ 0 & 1\end{bmatrix} \Rightarrow \boldsymbol{V^{-1}} = \begin{bmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1/4}\end{bmatrix}$$

Now, an expression for $y^T V^{-1} y$, will be:

$$y^T V^{-1} y = [Y_1 \quad Y_2]\begin{bmatrix}1 & 0\\ 0 & 1/4\end{bmatrix}\begin{bmatrix}Y_1\\ Y_2\end{bmatrix} = \begin{bmatrix}Y_1 & Y_2/4\end{bmatrix}\begin{bmatrix}Y_1\\ Y_2\end{bmatrix} \Rightarrow \boldsymbol{y^T V^{-1} y = Y_1^2 + \frac{Y_2^2}{4}}$$

As it was already shown above (in a), $Y_1^2 \sim \chi_1^2$. Now, it was also shown (in b) that $\left(\frac{Y_2-3}{2}\right)^2 \sim \chi_1^2$, and thus $\frac{Y_2^2}{4} \sim \chi_1^2$, plus a non-centrality parameter λ, which from property 6, is the following:

$$\boldsymbol{\lambda = \mu^T V^{-1} \mu} = [0 \quad 3]\begin{bmatrix}1 & 0\\ 0 & 1/4\end{bmatrix}\begin{bmatrix}0\\ 3\end{bmatrix} = \begin{bmatrix}0 & 3/4\end{bmatrix}\begin{bmatrix}0\\ 3\end{bmatrix} \Rightarrow \boldsymbol{\lambda = \frac{9}{4}}$$

And therefore, since we are adding two chi-squared distributed variables, with one degree of freedom each, it follows that (again from property 6):

$$\boldsymbol{y^T V^{-1} y = Y_1^2 + \frac{Y_2^2}{4} \sim \chi_2^2\left(\frac{9}{4}\right)}$$

Exercise 1.3:

Let the joint distribution of Y_1 and Y_2 be $N(\boldsymbol{\mu}, \boldsymbol{V})$ with:

$$\boldsymbol{\mu} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ and } \boldsymbol{V} = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}$$

a. Obtain an expression for $(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})$. What is its distribution?

b. Obtain an expression for $\boldsymbol{y}^T \boldsymbol{V}^{-1} \boldsymbol{y}$. What is its distribution?

SOLUTION:

A reminder from the book:

1.4.2 Chi-squared distribution:

1. The **central chi-squared distribution** with n degrees of freedom is defined as the sum of squares of n independent random variables $Z_1, \dots, Z_n$ each with the standard Normal distribution. It is denoted by:

$$X^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

In matrix notation, if $\boldsymbol{z} = [Z_1, \dots, Z_n]^T$ then $\boldsymbol{z}^T \boldsymbol{z} = \sum_{i=1}^{n} Z_i^2$ so that $X^2 = \boldsymbol{z}^T \boldsymbol{z} \sim \chi_n^2$.

2. If X^2 has the distribution χ_n^2, then its expected value is $E(X^2) = n$ and its variance is $Var(X^2) = 2n$.

3. If $Y_1, \dots, Y_n$ are independent Normally distributed random variables each with the distribution $Y_i \sim N(\mu_i, \sigma_i^2)$ then:

$$X^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu_i}{\sigma_i} \right)^2 \sim \chi_n^2$$

because each of the variables $Z_i = (Y_i - \mu_i)/\sigma_i$ has the standard Normal distribution $N(0, 1)$.

4. Let $Z_1, \dots, Z_n$ be independent random variables each with the distribution $N(0, 1)$ and let $Y_i = Z_i + \mu_i$, where at least one of the μ_i's is non-zero. Then the distribution of:

$$\sum Y_i^2 = \sum (Z_i + \mu_i)^2 = \sum Z_i^2 + 2\sum Z_i\mu_i + \sum \mu_i^2$$

has larger mean $n + \lambda$ and larger variance $2n + 4\lambda$ than χ_n^2 where $\lambda = \sum \mu_i^2$. This is called the **non-central chi-squared distribution** with n degrees of freedom and **non-centrality parameter** λ. It is denoted by $\chi_n^2(\lambda)$.

5. Suppose that the Y_i's are not necessarily independent and the vector $\boldsymbol{y} = [Y_1, \dots, Y_n]^T$ has the multivariate normal distribution $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$ where the variance-covariance matrix $\boldsymbol{V}$ is non-singular and its inverse is $\boldsymbol{V}^{-1}$. Then:

$$X^2 = (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \sim \chi_n^2$$

6. More generally if $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$ then the random variable $\boldsymbol{y}^T \boldsymbol{V}^{-1} \boldsymbol{y}$ has the non-central chi-squared distribution $\chi_n^2(\lambda)$ where $\lambda = \boldsymbol{\mu}^T \boldsymbol{V}^{-1} \boldsymbol{\mu}$.

7. If $X_1^2, \dots, X_m^2$ are m independent random variables with the chi-squared distributions $X_i^2 \sim \chi_{n_i}^2(\lambda_i)$, which may or may not be central, then their sum also has a chi-squared distribution with $\sum n_i$ degrees of freedom and non-centrality parameter $\sum \lambda_i$, i.e.,

$$\sum_{i=1}^{m} X_i^2 \sim \chi^2_{\sum_{i=1}^{m} n_i}\left(\sum_{i=1}^{m} \lambda_i\right)$$

This is called the **reproductive property** of the chi-squared distribution.

8. Let $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{y}$ has n elements but the Y_i's are not independent so that $\boldsymbol{V}$ is singular with rank $k < n$ and the inverse of $\boldsymbol{V}$ is not uniquely defined. Let $\boldsymbol{V}^-$ denote a generalized inverse of $\boldsymbol{V}$. Then the random variable $\boldsymbol{y}^T \boldsymbol{V}^- \boldsymbol{y}$ has the non-central chi-squared distribution with k degrees of freedom and non-centrality parameter $\lambda = \boldsymbol{\mu}^T \boldsymbol{V}^- \boldsymbol{\mu}$.

a. First, let us compute the inverse of Variance-Covariance matrix $\boldsymbol{V}$, $\boldsymbol{V}^{-1}$ as it will be needed. So:

$$\boldsymbol{V}^{-1} = \frac{1}{4 \cdot 9 - 1 \cdot 1}\begin{bmatrix} 9 & -1 \\ -1 & 4 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 9 & -1 \\ -1 & 4 \end{bmatrix} \Rightarrow \boldsymbol{V}^{-1} = \begin{bmatrix} \boldsymbol{9/35} & \boldsymbol{-1/35} \\ \boldsymbol{-1/35} & \boldsymbol{4/35} \end{bmatrix}$$

Since $y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ and $\mu = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, then their difference shall be:

$$y - \mu = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} Y_1 - 2 \\ Y_2 - 3 \end{bmatrix}$$

And therefore, the joint distribution, will have the following form:

$$(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) = [Y_1 - 2 \quad Y_2 - 3] \begin{bmatrix} {}^{9}/_{35} & {}^{-1}/_{35} \\ {}^{-1}/_{35} & {}^{4}/_{35} \end{bmatrix} \begin{bmatrix} Y_1 - 2 \\ Y_2 - 3 \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{9}{35}(Y_1 - 2) - \frac{1}{35}(Y_2 - 3) & -\frac{1}{35}(Y_1 - 2) + \frac{4}{35}(Y_2 - 3) \end{bmatrix} \begin{bmatrix} Y_1 - 2 \\ Y_2 - 3 \end{bmatrix} =$$

$$= \frac{9}{35}(Y_1 - 2)^2 - \frac{1}{35}(Y_2 - 3)(Y_1 - 2) - \frac{1}{35}(Y_1 - 2)(Y_2 - 3) + \frac{4}{35}(Y_2 - 3)^2 \Rightarrow$$

$$\Rightarrow (\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) = \frac{\mathbf{9}}{\mathbf{35}}(\boldsymbol{Y_1} - \mathbf{2})^2 - \frac{\mathbf{2}}{\mathbf{35}}(\boldsymbol{Y_1} - \mathbf{2})(\boldsymbol{Y_2} - \mathbf{3}) + \frac{\mathbf{4}}{\mathbf{35}}(\boldsymbol{Y_2} - \mathbf{3})^2$$

From property 5, we know that for a multivariate normal distribution $y \sim N(\mu, V)$, the quadratic form $(y - \mu)^T V^{-1} (y - \mu)$ follows a chi-squared distribution with degrees of freedom equal to the dimension of y (and in this case, we have only two dimensions), and therefore:

$$(\boldsymbol{y} - \boldsymbol{\mu})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) = \frac{\mathbf{9}}{\mathbf{35}}(\boldsymbol{Y_1} - \mathbf{2})^2 - \frac{\mathbf{2}}{\mathbf{35}}(\boldsymbol{Y_1} - \mathbf{2})(\boldsymbol{Y_2} - \mathbf{3}) + \frac{\mathbf{4}}{\mathbf{35}}(\boldsymbol{Y_2} - \mathbf{3})^2 \sim \boldsymbol{\chi}_{\mathbf{2}}^{\mathbf{2}}$$

b. From the previous question (a), we already know that:

$$V^{-1} = \begin{bmatrix} {}^{9}/_{35} & {}^{-1}/_{35} \\ {}^{-1}/_{35} & {}^{4}/_{35} \end{bmatrix}$$

And therefore, the expression for $y^T V^{-1} y$ shall be:

$$y^T V^{-1} y = [Y_1 \quad Y_2] \begin{bmatrix} {}^{9}/_{35} & {}^{-1}/_{35} \\ {}^{-1}/_{35} & {}^{4}/_{35} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \frac{9}{35} \cdot Y_1 - \frac{1}{35} \cdot Y_2 & -\frac{1}{35} \cdot Y_1 + \frac{4}{35} \cdot Y_2 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} =$$

$$= \frac{9}{35} \cdot Y_1^2 - \frac{1}{35} \cdot Y_1 \cdot Y_2 - \frac{1}{35} \cdot Y_2 \cdot Y_1 + \frac{4}{35} \cdot Y_2^2 \Rightarrow$$

$$\Rightarrow \boldsymbol{y^T V^{-1} y = \frac{9}{35} \cdot Y_1^2 - \frac{2}{35} \cdot Y_1 \cdot Y_2 + \frac{4}{35} \cdot Y_2^2}$$

Now, the distribution of $y^T V^{-1} y$ is a more general case of the one described in the previous question (a) and thus follows property 6, meaning that: "if $\boldsymbol{y} \sim \boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{V})$ then the random variable $\boldsymbol{y}^T \boldsymbol{V}^{-1} \boldsymbol{y}$ has the non-central chi-squared distribution $\chi_n^2(\lambda)$ where $\lambda = \boldsymbol{\mu}^T \boldsymbol{V}^{-1} \boldsymbol{\mu}$."

In our case, y can be written as $y = \mu + Z$, where: $Z \sim N(0, V)$. So if we expanded on this, we would have:

$$y^T V^{-1} y = (\mu + Z)^T V^{-1} (\mu + Z) = \mu^T V^{-1} \mu + 2\mu^T V^{-1} Z + Z^T V^{-1} Z$$

with:

- $Z^T V^{-1} Z \sim \chi_2^2$, because $Z \sim N(0, V)$, and the quadratic form of a multivariate normal distribution follows a chi-squared distribution with degrees of freedom equal to the dimension of Z (which is 2).
- $2\mu^T V^{-1} Z$ is normally distributed with a mean of 0.

Thus $y^T V^{-1} y$ is a sum of a chi-squared distribution and a normal distribution. This means that $y^T V^{-1} y$ follows a non-central chi-squared distribution with 2 degrees of freedom and a non-centrality parameter $\lambda = \mu^T V^{-1} \mu$, which is:

$$\lambda = \mu^T V^{-1} \mu = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} 9/35 & -1/35 \\ -1/35 & 4/35 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{18}{35} - \frac{3}{35} & -\frac{2}{35} + \frac{12}{35} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{15}{35} & \frac{10}{35} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \frac{30}{35} + \frac{30}{35} = \frac{60}{35} \Rightarrow \boldsymbol{\lambda = \frac{12}{7}}$$

Therefore, in conclusion:

$$\boldsymbol{y^T V^{-1} y = \frac{9}{35} \cdot Y_1^2 - \frac{2}{35} \cdot Y_1 \cdot Y_2 + \frac{4}{35} \cdot Y_2^2 \sim \chi_2^2\left(\frac{12}{7}\right)}$$

Exercise 1.4:

Let $\boldsymbol{Y_1, \dots, Y_n}$ be independent random variables each with the distribution $\boldsymbol{N(\mu, \sigma^2)}$. Let:

$$\boldsymbol{\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i} \quad \textbf{and} \quad \boldsymbol{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.}$$

a. What is the distribution of $\boldsymbol{\bar{Y}}$?

b. Show that $\boldsymbol{S^2 = \frac{1}{n-1}[\sum_{i=1}^{n}(Y_i - \mu)^2 - n(\bar{Y} - \mu)^2]}$.

c. From (b) it follows that $\boldsymbol{\sum (Y_i - \mu)^2/\sigma^2 = (n-1)\,S^2/\sigma^2 + [(\bar{Y} - \mu)^2 n/\sigma^2]}$. How does this allow you to deduce that $\boldsymbol{\bar{Y}}$ and $\boldsymbol{S^2}$ are independent?

d. What is the distribution of $\boldsymbol{\frac{(n-1)S^2}{\sigma^2}}$?

e. What is the distribution of $\boldsymbol{\frac{\bar{Y}-\mu}{S/\sqrt{n}}}$?

SOLUTION:

a. Since the Y_i are independent and each has the distribution $N(\mu, \sigma^2)$, the expectation of $\bar{Y}$ is:

$$E(\bar{Y}) = E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i] = \frac{1}{n} \cdot n \cdot \mu = \mu$$

while its variance is:

$$Var(\bar{Y}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

We know that $\bar{Y}$ consists of a linear combination of independent, normally distributed variables and therefore it is itself normally distributed. Thus, the distribution of $\bar{Y}$ shall be:

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

b. Let us start from the definition of the sample variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}[(Y_i - \mu) - (\bar{Y} - \mu)]^2 =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}[(Y_i-\mu)^2 - 2\cdot(Y_i-\mu)(\bar{Y}-\mu) + (\bar{Y}-\mu)^2] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - \sum_{i=1}^{n}2\cdot(Y_i-\mu)(\bar{Y}-\mu) + \sum_{i=1}^{n}(\bar{Y}-\mu)^2\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - 2\cdot(\bar{Y}-\mu)\sum_{i=1}^{n}(Y_i-\mu) + \sum_{i=1}^{n}(\bar{Y}-\mu)^2\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - 2\cdot(\bar{Y}-\mu)\cdot n\cdot(\bar{Y}-\mu) + n\cdot(\bar{Y}-\mu)^2\right] =$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - 2\cdot n\cdot(\bar{Y}-\mu)^2 + n\cdot(\bar{Y}-\mu)^2\right] \Rightarrow$$

$$\Rightarrow \boldsymbol{S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(Y_i-\mu)^2 - n\cdot(\bar{Y}-\mu)^2\right]}$$

c. We are given the following expression:

$$\frac{\sum_{i=1}^{n}(Y_i-\mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{Y}-\mu)^2}{\sigma^2}$$

So, why are $\bar{Y}$ and S^2 independent? Let us look at the two right hand terms one by one.

Firstly, let us discuss the term:

$$\frac{(n-1)S^2}{\sigma^2}$$

Here S^2 is the sample variance, which is defined as:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

Therefore:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{n-1}{\sigma^2}\cdot\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

The sample variance measures the spread of the individual Y_i's around the sample mean $\bar{Y}$. This involves $n-1$ degrees of freedom because the calculation of S^2 depends on n data points, but the sample mean $\bar{Y}$ is used to estimate the center of the data, reducing the degrees of freedom by 1.

Thus, under the assumption that the Y_i's are normally distributed, the sum of squares, which was defined above, follows a chi-squared distribution with $n-1$ degrees of freedom:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Secondly, let us discuss the term:

$$\frac{n(\bar{Y}-\mu)^2}{\sigma^2}$$

And from question **a**, we already know that:

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Therefore:

$$\frac{n(\bar{Y}-\mu)^2}{\sigma^2} = \left(\frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim Z^2$$

where Z is a standard normal random variable, $Z \sim N(0,1)$. And hence:

$$\frac{n(\bar{Y}-\mu)^2}{\sigma^2} \sim \chi^2_1$$

Since the total sum of squares can be split into two independent components, one involving $\bar{Y}$ and the other involving S^2, then by Cochran's Theorem, the chi-squared terms must be independent. More formally, the independence of χ^2_1 and χ^2_{n-1} implies that **$\bar{Y}$ and S^2 are independent**.

Cochran's Theorem:

Cochran's Theorem provides a way to decompose sums of squared normal random variables into independent chi-squared distributions. Specifically, if you have a set of independent normal random variables $Y_1, Y_2, \ldots, Y_n$ drawn from $N(\mu, \sigma^2)$. In our example, Cochran's Theorem states that the total sum of squares:

$$\sum_{i=1}^{n}(Y_i - \mu)^2$$

can be decomposed into two independent components:

$$\frac{n(\bar{Y} - \mu)^2}{\sigma^2} \sim \chi_1^2$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

This result is a key property of normal distributions and is a consequence of the fact that the sample mean $\bar{Y}$ and sample variance S^2 capture independent aspects of the data. $\bar{Y}$ captures location (center), while S^2 captures spread (variability) around the center.

d. As it was already shown in question **c**:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

e. From question **c**, we got that:

$$\frac{n(\bar{Y} - \mu)^2}{\sigma^2} = \left(\frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim Z^2 \Longrightarrow \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 \Longrightarrow S \sim \frac{\sigma}{\sqrt{n-1}} \cdot \sqrt{\chi_{n-1}^2}$$

Thus, the numerator follows a standard normal distribution, while the denominator involves the sample standard deviation, which is related to the chi-squared distribution with $n-1$ degrees of freedom. So, when we take the ratio of a standard normal random variable and the square root of a chi-squared random variable (divided by its degrees of freedom), the result follows a **t-distribution**. Hence:

$$T = \frac{\bar{Y} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Exercise 1.5:

This exercise is a continuation of the example in Section 1.6.2 in which $\boldsymbol{Y_1, \ldots, Y_n}$ are independent Poisson random variables with the parameter $\boldsymbol{\theta}$.

a. Show that $\boldsymbol{E(Y_i) = \theta}$ for $\boldsymbol{i = 1, \ldots, n}$.

b. Suppose $\boldsymbol{\theta = e^{\beta}}$. Find the maximum likelihood estimator of $\boldsymbol{\beta}$.

c. Minimize $\boldsymbol{S = \sum(Y_i - e^{\beta})^2}$ to obtain a least squares estimator of $\boldsymbol{\beta}$.

SOLUTION:

a. We are given that $Y_1, \ldots, Y_n$ are independent Poisson random variables with the parameter θ. Therefore:

$$E(Y_i) = \sum_{k=0}^{\infty} k \cdot P(Y_i = k) = \sum_{k=0}^{\infty} k \cdot \frac{\theta^k \cdot e^{-\theta}}{k!}$$

However, when $k = 0$, the whole term becomes zero, thus it is superfluous in our expression. We can take it out:

$$E(Y_i) = \sum_{k=1}^{\infty} k \cdot \frac{\theta^k \cdot e^{-\theta}}{k!} = \sum_{k=1}^{\infty} \frac{\theta^k \cdot e^{-\theta}}{(k-1)!} = e^{-\theta} \sum_{k=1}^{\infty} \frac{\theta^k}{(k-1)!} = e^{-\theta} \sum_{k=1}^{\infty} \frac{\theta \cdot \theta^{k-1}}{(k-1)!} \Longrightarrow$$

$$\Longrightarrow E(Y_i) = \theta \cdot e^{-\theta} \sum_{k=1}^{\infty} \frac{\theta^{k-1}}{(k-1)!}$$

Setting $j = k - 1$, we get:

$$E(Y_i) = \theta \cdot e^{-\theta} \sum_{j=0}^{\infty} \frac{\theta^j}{j!} = \theta \cdot e^{-\theta} \cdot e^{\theta} \Longrightarrow \boldsymbol{E(Y_i) = \theta}$$

b. Given that $Y_1, \ldots, Y_n$ are independent Poisson random variables with parameter $\theta = e^{\beta}$, the probability mass function for each Y_i is:

$$P(Y_i = y_i) = \frac{\left(e^{\beta}\right)^{y_i} \cdot e^{-e^{\beta}}}{y_i!}$$

The likelihood function $L(\beta)$ is the product of the individual probabilities for all Y_i's:

$$L(\beta) = \prod_{i=1}^{n} P(Y_i = y_i) = \prod_{i=1}^{n} \frac{\left(e^{\beta}\right)^{y_i} \cdot e^{-e^{\beta}}}{y_i!} = \frac{\left(e^{\beta}\right)^{\sum_{i=1}^{n} y_i} \cdot e^{-n \cdot e^{\beta}}}{\prod_{i=1}^{n} y_i!}$$

The log-likelihood function $l(\beta)$ is the natural logarithm of the likelihood function:

$$l(\beta) = lnL(\beta) = ln\left(\frac{\left(e^{\beta}\right)^{\sum_{i=1}^{n} y_i} \cdot e^{-n \cdot e^{\beta}}}{\prod_{i=1}^{n} y_i!}\right) = ln\left[\left(e^{\beta}\right)^{\sum_{i=1}^{n} y_i}\right] + ln\left[e^{-n \cdot e^{\beta}}\right] - ln\left[\prod_{i=1}^{n} y_i!\right] \Rightarrow$$

$$\Rightarrow l(\beta) = \beta \cdot \sum_{i=1}^{n} y_i - n \cdot e^{\beta} - ln\left[\prod_{i=1}^{n} y_i!\right]$$

To find the maximum likelihood estimator of β, we take the derivative of $l(\beta)$ with respect to β and set it equal to zero:

$$\frac{d}{d\beta} l(\beta) = 0 \Rightarrow \frac{d}{d\beta}\left(\beta \cdot \sum_{i=1}^{n} y_i\right) - \frac{d}{d\beta}\left(n \cdot e^{\beta}\right) - \frac{d}{d\beta}\left(ln\left[\prod_{i=1}^{n} y_i!\right]\right) = 0 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} y_i - n \cdot e^{\beta} - 0 = 0 \Rightarrow e^{\beta} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i \Rightarrow e^{\beta} = \bar{Y} \Rightarrow ln\left(e^{\beta}\right) = ln(\bar{Y}) \Rightarrow$$

$$\Rightarrow \boldsymbol{\beta} = \boldsymbol{ln}(\bar{\boldsymbol{Y}})$$