Quadratic Forms uniformly used and analytically solvable.

# Chapter 2. Distributions and Quadratic Forms

## 2.1 Random Vector

(a) Expectation: The expected value of a random vector

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \text{ is given by } \boldsymbol{\mu} \equiv E(\boldsymbol{Y}) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix}.$$

(b)Covariance Matrix:

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \; Cov(\boldsymbol{Y}) = E\{[\boldsymbol{Y} - E(\boldsymbol{Y})][\boldsymbol{Y} - E(\boldsymbol{Y})]^\top\} = E(\boldsymbol{Y}\boldsymbol{Y}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top.$$

-Let $\boldsymbol{A}$ be a constant matrix, then

$$Cov(\boldsymbol{AY}) = \boldsymbol{A}[Cov(\boldsymbol{Y})]\boldsymbol{A}^\top.$$

-Let $\boldsymbol{A}, \boldsymbol{B}$ be constant matrices, then

$Cov(\boldsymbol{AY}, \boldsymbol{BY}) = \boldsymbol{A}Cov($[redacted]$\boldsymbol{Y})\boldsymbol{B}^\top.$

(c)Trace of a matrix:

The theorems in these two chapters make considerable use of the trace of a matrix. The trace of a matrix is the sum of its diagonal elements. The important properties of the trace of a matrix include the following:

1. It is the sum of its eigenvalues.

2. It is equal to the rank of an idempotent matrix.

3. Products are cyclically commutative, for example, If any terms below are well defined.

   cyclic permutation, called also.

$$tr(\boldsymbol{ABC}) = tr(\boldsymbol{BCA}) = tr(\boldsymbol{CAB}).$$

Note: they have cyclic order!

4. For a quadratic form, we have

$$\boldsymbol{Y}^\top\boldsymbol{AY} = tr(\boldsymbol{Y}^\top\boldsymbol{AY}) = tr(\boldsymbol{AYY}^\top).$$

random vector.

(d) Symmetric matrices

For two vectors $\boldsymbol{x}, \boldsymbol{y}$, an expression of the form $\boldsymbol{x}^\top\boldsymbol{Ay}$ is called a *bilinear form*. For example,

$$\boldsymbol{x}^\top\boldsymbol{Ay} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = 4x_1y_1 + 8x_1y_2 - 2x_2y_1 + 7x_2y_2.$$

any matrix.

When $\boldsymbol{x}$ is used in place of $\boldsymbol{y}$, the expression becomes $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$. It is then called a *quadratic form* and is a quadratic function of $x's$. Then, we have

same

$$\begin{aligned}
\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 8 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= 4x_1^2 + (8-2)x_1x_2 + 7x_2^2 \\
&= 4x_1^2 + (3+3)x_1x_2 + 7x_2^2 \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.
\end{aligned}$$

In this way, we can write any quadratic form $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ as $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x}$ where $\boldsymbol{B} = \frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^\top)$ is symmetric. While we can write every quadratic form as $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ for an infinite number of matrices, we can only write $\boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x}$ one way for $\boldsymbol{B}$ symmetric. For example,

$$4x_1^2 + 6x_1x_2 + 7x_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 3+a \\ 3-a & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

for any value of $a$. However, the matrix is symmetric only when $a = 0$. This means that for any particular quadratic form, there is only one unique matrix such that the quadratic form can be written as $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A}$ symmetric. Because of the uniqueness of this symmetric matrix, all further discussion of quadratic form $\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$ is confined to the case of $A$ being symmetric. In particular, for a vector $\boldsymbol{y} = (y_1, \ldots, y_p)^\top$ and a symmetric matrix $\boldsymbol{A}_{p\times p} = [a_{ij}]$,

$$\boldsymbol{y}^\top \boldsymbol{A}\boldsymbol{y} = \sum_{j=1}^{p}\sum_{i=1}^{p} a_{ij}y_iy_j = \sum_{i=1}^{p} a_{ii}y_i^2 + \sum_{i\neq j} a_{ij}y_iy_j. \quad \binom{p}{2}$$

dominating the complexity.

*Remark:* Note that $\boldsymbol{Y}^\top \boldsymbol{A}\boldsymbol{Y}$ is not linear function of $\boldsymbol{Y}$, $E(\boldsymbol{Y}^\top \boldsymbol{A}\boldsymbol{Y}) \neq E(\boldsymbol{Y}^\top)\boldsymbol{A}E(\boldsymbol{Y})$ in general.

**THEOREM 1.** *Let $\boldsymbol{Y}$ be a random vector with mean $\boldsymbol{\mu} = E(\boldsymbol{Y})$ and $\boldsymbol{\Sigma} = Cov(\boldsymbol{Y})$. Then, $E(\boldsymbol{Y}^\top \boldsymbol{A}\boldsymbol{Y}) = \boxed{tr(\boldsymbol{A}\boldsymbol{\Sigma})} + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$.*

**Proof:** First, by the definition of covariance matrix of a random vector,

$$\boldsymbol{\Sigma} = E(\boldsymbol{Y}\boldsymbol{Y}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top \Rightarrow E(\boldsymbol{Y}\boldsymbol{Y}^\top) = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top.$$

Next, since $\boldsymbol{Y}^\top \boldsymbol{A}\boldsymbol{Y}$ is a scalar, it equals to its trace. Thus,

compute the mean.

$$\begin{aligned} E(\boldsymbol{Y}^\top \boldsymbol{A}\boldsymbol{Y}) &= E(tr(\boldsymbol{Y}^\top \boldsymbol{A}\boldsymbol{Y})) = E(tr(\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^\top)) \\ &= tr(E(\boldsymbol{A}\boldsymbol{Y}\boldsymbol{Y}^\top)) \\ &= tr(\boldsymbol{A}E(\boldsymbol{Y}\boldsymbol{Y}^\top)) \\ &= tr(\boldsymbol{A}(\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top)) \\ &= tr(\boldsymbol{A}\boldsymbol{\Sigma} + \boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^\top) \\ &= tr(\boldsymbol{A}\boldsymbol{\Sigma}) + tr(\boldsymbol{A}\boldsymbol{\mu}\boldsymbol{\mu}^\top) \\ &= tr(\boldsymbol{A}\boldsymbol{\Sigma}) + tr(\boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}) \\ &= tr(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}. \end{aligned}$$

The proof is complete.

-Moment generating function (M.G.F): the moment generating function of a random vector $\boldsymbol{Y}$ is given by

$$M_{\boldsymbol{Y}}(t) = E(e^{t^\top \boldsymbol{Y}}),$$

Strong assumption : Y has infinite moments.

where $\boldsymbol{t} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{pmatrix}$ if the expectation exists for $-h < t_i < h$ where $h > 0$ and $i = 1, \cdots, n$.

**THEOREM 2.** *Let $g_1(\boldsymbol{Y_1}), \cdots, g_m(\boldsymbol{Y_m})$ be $m$ functions of the random vectors $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_m$ respectively. If $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_m$ are mutually independent, then $g_1, \cdots, g_m$ are mutually independent.*

verify a distribution

1. pdf 2. cdf 3. MGF 4. chf

**2.2 Multivariate Normal Distribution**

(a) Probability density function (p.d.f) of $\boldsymbol{Y_{p\times 1}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

normalizing the high dimension integration.

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}}(2\pi)^{-\frac{p}{2}}e^{-\frac{1}{2}[(\boldsymbol{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})]}.$$

(b) Moment generating function of $\boldsymbol{Y_{p\times 1}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$M_{\boldsymbol{Y}}(\boldsymbol{t}) = e^{\boldsymbol{t}^\top \boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top \boldsymbol{\Sigma}\boldsymbol{t}}.$$

-Let $\boldsymbol{B}$ be a constant matrix and $\boldsymbol{c}$ be a constant vector.

$$\boldsymbol{B}\boldsymbol{Y} + \boldsymbol{c} \sim N(\boldsymbol{B}\boldsymbol{\mu} + \boldsymbol{c}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^\top).$$

(c) Marginal distribution, Conditional Distribution and independence.

Let

$|\Sigma|^{-\frac{1}{2}}(2\pi)^{-\frac{p}{2}} \exp\{\frac{-1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)\}$.

Countor? SIR? What's it?

VAE : transform unknow dist to $\mathcal{N}(0,1)$ and recover the dist by neural networks.



$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} \sim N\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right],$$

then

conditional distribution & projection.

1. $\boldsymbol{Y_1} \sim N(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_{11}})$.

2. $\boldsymbol{Y_1}|\boldsymbol{Y_2} = \boldsymbol{y_2} \sim N(\boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}(\boldsymbol{y_2} - \boldsymbol{\mu_2}), \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}^{-1}}\boldsymbol{\Sigma_{21}})$. Deduce it by myself. □

3. $\boldsymbol{Y_1}$ and $\boldsymbol{Y}_2$ are independent iff $\boldsymbol{\Sigma}_{12} = 0$.

-Non-Central $\chi^2$ distribution

It is known that the density function of $u \sim \chi^2_{(n)}$, a central $\chi^2$ distribution, is

$$f(u) = \frac{u^{\frac{1}{2}n-1}e^{-\frac{1}{2}u}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \qquad (1)$$

← No need to remember.

Central $\chi^2$ distribution.

-Let $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I_n})$, then $\boldsymbol{x}^\top \boldsymbol{x} \sim \chi^2_{(n)}$.

-Let $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{I_n})$, then $u = \boldsymbol{x}^\top \boldsymbol{x} \sim \chi^2_{(n,\lambda)}$, where $\boxed{\lambda = \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{\mu}}$ is a non-centered parameter and the density function of $\chi^2_{(n,\lambda)}$ is

indepence is important

$$f(u) = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\frac{u^{\frac{1}{2}n+k-1}e^{-\frac{1}{2}u}}{2^{\frac{1}{2}n+k}\Gamma(\frac{1}{2}n+k)}, \quad u > 0, \lambda \geq 0. \qquad (2)$$

Define $\lambda^k = 1$ when $\lambda = 0$ and $k = 0$.

-The moment generating function of $u \sim \chi^2_{(n,\lambda)}$ is

$$\boxed{(1-2t)^{-\frac{n}{2}}}\boxed{e^{-\lambda[1-(1-2t)^{-1}]}}.$$

*Remark:* When $\lambda = 0$, the above M.G.F is $\boxed{(1-2t)^{-\frac{n}{2}}}$ which is precisely the M.G.F of $\chi^2_{(n)}$.

-Non-Central $F$ distribution.

the denumerators should be central $\chi^2$

Let $u_1 \sim \chi^2_{(p_1,\lambda)}$ and $u_2 \sim \chi^2_{\cancel{(p_2),0}}$ $(p_2, 0)$. And $u_1$ is independent of $u_2$. Then,

the numerators are non-central

$$w = \frac{u_1/p_1}{u_2/p_2} \sim F_{(p_1,p_2,\lambda)}. \qquad (3)$$

Let $z \sim N(\mu, 1)$ and $u \sim \chi^2_{(n)}$. And $z$ is independent of $u$. Then,

$$t = \frac{z}{\sqrt{\frac{u}{n}}} \sim \text{Non-central } t \text{ distribution.}$$