Matrix Algebra and Some Distribution Theory

Economics 620 - Spring 1999

Professor N.M.Kiefer T.A.: Chang Park

1. Matrix Algebra

1.1 Partitioned matrix and its inverse

- Addition and multiplication

$$A+B = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} \\ A_{21}+B_{21} & A_{22}+B_{22} \end{bmatrix}$$

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22} \\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{bmatrix}$$

- Determinant of a partitioned matrix

$$\begin{aligned} |A| &= \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \\ &= |A_{22}| \left| A_{11} - A_{12}A_{22}^{-1}A_{21} \right| \\ &= |A_{11}| \left| A_{22} - A_{21}A_{11}^{-1}A_{12} \right| \end{aligned}$$

  provided that both A_{11} and A_{22} are non-singular.
- Inverse of a partitioned matrix

$$\begin{aligned} A^{-1} &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} F_{11} & -F_{11}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}F_{11} & A_{22}^{-1}+A_{22}^{-1}A_{21}F_{11}A_{12}A_{22}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}F_{22}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}F_{22} \\ -F_{22}A_{21}A_{11}^{-1} & F_{22} \end{bmatrix} \end{aligned}$$

  where $F_{11} = \left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}$ and $F_{22} = \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1}$ provided that both A_{11} and A_{22} are non-singular.
- An important application of inverse of partitioned matrix; Suppose we partition a matrix X as $\underset{(n\times K)}{X} = \left[\underset{(n\times l)}{X_1} \quad \underset{(n\times(K-l))}{X_2}\right]$:

$$X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} \left(X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1\right)^{-1} & \cdots \\ \cdots & \cdots \end{bmatrix}$$

- The upper left block $(l \times l)$ matrix can be expressed as

$$\begin{aligned} \left(X_1'X_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1\right)^{-1} &= \left[X_1'\left(I - X_2(X_2'X_2)^{-1}X_2'\right)X_1\right]^{-1} \\ &= [X_1'M_2X_1]^{-1} \end{aligned}$$

  where $M_2 = \left(I - X_2(X_2'X_2)^{-1}X_2'\right)$. The matrix M_2 plays quite an important and unique role in multiple regression model.

- M_2 is symmetric and idempotent.

$$\begin{aligned} M_2M_2' &= \left(I - X_2(X_2'X_2)^{-1}X_2'\right)\left(I - X_2(X_2'X_2)^{-1}X_2'\right)' \\ &= I - X_2(X_2'X_2)^{-1}X_2' - X_2(X_2'X_2)^{-1}X_2' + X_2(X_2'X_2)^{-1}X_2'X_2(X_2'X_2)^{-1}X_2' \\ &= I - X_2(X_2'X_2)^{-1}X_2' = M_2 \end{aligned}$$

- $M_2X_2 = 0$.

1.2 Eigenvectors and eigenvalues of real matrix

Suppose the solution of the following system of equations;

$$Ax = \lambda x$$

where A is an $(n \times n)$ square matrix, x is a non null $(n \times 1)$ vector and λ is a scalar. The λ's satisfying the system of equations are called eigenvalues(characteristic values, latent values) and x's eigenvectors(characteristic vectors, latent vectors). We can rewrite the system as

$$(A - \lambda I)x = 0$$

If the matrix $(A - \lambda I)$ is non-singular - i.e. its inverse exists, the only solution is the trivial solution, $x = 0$. In order for a non-trivial solution to exist, we should have a singular matrix $(A - \lambda I)$; which implies that

$$|A - \lambda I| = 0$$

We can find the eigenvalues of a matrix A by expanding above determinant and solving the nth order polymonial equation.

Example;

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

Then,

$$A - \lambda I = \begin{bmatrix} 4 - \lambda & 2 \\ 2 & 1 - \lambda \end{bmatrix}$$

Therefore,

$$|A - \lambda I| = (4 - \lambda)(1 - \lambda) - 4 = 0$$

The solutions for the equation is given by;

$$\lambda_1 = 5 \text{ and } \lambda_2 = 0$$

For $\lambda_1 = 5$;

$$\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \Rightarrow x_1 = 2x_2$$

One element in the eigenvector is always arbitrary. The usual practice is to normalize the vector so that we have unit length for the eigenvectors, i.e., $x_1^2 + x_2^2 = 1$. Then, the eigenvector corresponding to $\lambda_1 = 5$ is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

For $\lambda_2 = 0$, it is given by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{bmatrix}$$

Note that the matrix X whose columns consist of eigenvectors of a matrix is an orthogonal matrix - columns of the matrix are orthogonal-;

$$X = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix}$$

Moreover,

$$X'X = XX' = I$$

and

$$X^{-1} = X'$$

Suppose X is $(n \times n)$ real matrix.

- **The sum of the eigenvalues of X is equal to the sum of its diagonal elements(trace).**
- The product of the eigenvalues of X is equal to its determinants.
- **The rank of X is equal to the number of its non-zero eigenvalues.**
- The eigenvalues of X^{-1} are the reciprocals of those of X, but eigenvectors are the same.
- **Each eigenvalues of an idempotent matrix is either 0 or 1.**
- **The rank of an idempotent matrix is equal to its trace.**

Now, we assume that X is $(n \times n)$ **symmetric** matrix as well as real.

- The eigenvalues of X are real.
- Eigenvectors corresponding to distinct eigenvalues are pairwise orthogonal.
- **The orthogonal matrix of eigenvectors diagonalizes X**, i.e., $X = \Gamma' \Lambda \Gamma$ where Γ is $(n \times n)$ orthogonal matrix whose columns consist of eigenvectors of X and Λ is $(n \times n)$ diagonal matrix whose main diagonals consist of eigenvalues of X.
- Any *symmetric positive definite* matrix X can be factored into LL' where L is a lower

triangular matrix. It is called the *Cholesky* decomposition.

1.3 Rank of a matrix

- The maximum number of linearly independent rows is equal to the maximum number of linearly independent columns. This number is the rank of the matrix, denoted by $\rho(X)$.
- $\rho(X) \leq \min(m, n)$ where m and n are row and column dimensions of a matrix X.
- $\rho(X) = \rho(X')$.
- If $\rho(X) = m = n$, X is non-singular and a unique inverse X^{-1} exists.
- $\rho(X'X) = \rho(XX') = \rho(X)$.
- If P and Q are non-singular matrices of orders m and n, then $\rho(PX) = \rho(XQ) = \rho(PXQ) = \rho(X)$.
- $\rho(XY) \leq \min[\rho(X), \rho(Y)]$ where Y is $(n \times l)$ matrix.

1.4 Kronecker products

- The Kronecker product is defined as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1K}B \\ a_{21}B & a_{22}B & \cdots & a_{2K}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nK}B \end{bmatrix}$$

- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ if A and B are square and non-singular.
- $(A \otimes B)' = A' \otimes B'$.
- If A is $(l \times l)$ and B is $(n \times n)$, then $|A \otimes B| = |A|^l |B|^n$.
- $tr(A \otimes B) = tr(A)\, tr(B)$.
- $(A \otimes B)(C \otimes D) = AC \otimes BD$.
- $A \otimes (B + C) = A \otimes B + A \otimes C$. $A \otimes (B \otimes C) = (A \otimes B) \otimes C$.

1.5 vec and vech operators

- Suppose that X is $(m \times n)$ matrix expressed as $X = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}$ where X_i is the ith column of the matrix X. Then

$$\underset{(mn \times 1)}{vec(X)} = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix}$$

- $vec(ABC) = (C' \otimes A)\, vec(B)$.
- Suppose that X is $(n \times n)$ matrix. $vech$ operator transforms an $(n \times n)$ matrix into

an $\left[\frac{n(n+1)}{2} \times 1\right]$ vector by vertically stacking those elements on or below the main diagonal. For example,

$$\text{vech}\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ a_{22} \\ a_{32} \\ a_{33} \end{bmatrix}$$

1.6 Matrix Differentiation

- Let a and x are $(k \times 1)$ vectors and A is an $(k \times k)$ matrix.

$$\begin{aligned} \frac{\partial (a'x)}{\partial x} &= a \qquad \frac{\partial (a'x)}{\partial x'} = a' \\ \frac{\partial (x'Ax)}{\partial x} &= (A + A')\,x \qquad \frac{\partial (x'Ax)}{\partial x \partial x'} = (A + A') \\ \frac{\partial (x'Ax)}{\partial A} &= xx' \end{aligned}$$

- We don't want to prove the claim rigorously. But

$$a'x = \sum_{i=1} a_i x_i$$

If you want to differentiate the function with respect to x, you have to differentiate the function with respect to each element of vector x and form a vector -called gradient- with the result.

$$\frac{\partial (a'x)}{\partial x} = \begin{bmatrix} \frac{\partial (a'x)}{\partial x_1} \\ \frac{\partial (a'x)}{\partial x_2} \\ \cdots \\ \frac{\partial (a'x)}{\partial x_k} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_k \end{bmatrix} = a$$

You can understand $\frac{\partial (a'x)}{\partial x'}$ simply as the transpose of $\frac{\partial (a'x)}{\partial x}$: For the differentiation of the quadratic form, consider the summation expression;

$$\begin{aligned} x'Ax &= \sum_{i=1} \sum_{j=1} x_i a_{ij} x_j \\ &= x_1 a_{11} x_1 + x_1 a_{12} x_2 + x_1 a_{13} x_3 + \cdots + x_1 a_{1k} x_k \\ &\quad + x_2 a_{21} x_1 + x_2 a_{22} x_2 + x_2 a_{23} x_3 + \cdots + x_2 a_{2k} x_k \\ &\quad + x_3 a_{31} x_1 + x_3 a_{32} x_2 + x_3 a_{33} x_3 + \cdots + x_3 a_{3k} x_k \\ &\quad + \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\quad + x_k a_{k1} x_1 + x_k a_{k2} x_2 + x_k a_{k3} x_3 + \cdots + x_k a_{kk} x_k \end{aligned}$$

Now, we have

$$\begin{aligned}
\frac{\partial (x'Ax)}{\partial x_1} &= 2a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1k}x_k \\
&\quad + x_2a_{21} + x_3a_{31} + \cdots + x_ka_{k1} \\
&= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1k}x_k \\
&\quad + a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + \cdots + a_{k1}x_k \\
&= A_1x + A^{1\prime}x = \left(A_1 + A^{1\prime}\right)x
\end{aligned}$$

where A_1 is the first row of the matrix A and A^1 is the first column of the matrix A. Similarly,

$$\begin{aligned}
\frac{\partial (x'Ax)}{\partial x_2} &= a_{21}x_1 + 2a_{22}x_2 + a_{23}x_3 + \cdots + a_{2k}x_k \\
&\quad + x_1a_{12} + x_3a_{32} + \cdots + x_ka_{k2} \\
&= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2k}x_k \\
&\quad + a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + \cdots + a_{k2}x_k \\
&= A_2x + A^{2\prime}x = \left(A_2 + A^{2\prime}\right)x
\end{aligned}$$

You see the pattern emerging from the calculation. In general,

$$\frac{\partial (x'Ax)}{\partial x_i} = \left(A_i + A^{i\prime}\right)x \qquad i = 1, 2, \cdots, k$$

We stack the vectors to get;

$$\frac{\partial (x'Ax)}{\partial x} = \begin{bmatrix} \frac{\partial(x'Ax)}{\partial x_1} \\ \frac{\partial(x'Ax)}{\partial x_2} \\ \cdots \\ \frac{\partial(x'Ax)}{\partial x_k} \end{bmatrix} = \left[\begin{bmatrix} A_1 \\ A_2 \\ \cdots \\ A_k \end{bmatrix} + \begin{bmatrix} A^{1\prime} \\ A^{2\prime} \\ \cdots \\ A^{k\prime} \end{bmatrix}\right] x = (A + A')\,x$$

You can verify the result for $\frac{\partial(x'Ax)}{\partial A} = xx'$ with a similar argument.

- Consider the following minimization problem;

$$\begin{aligned}
\min_b S(b) &= \min_b (y - Xb)'(y - Xb) \\
&= \min_b y'y - y'Xb - b'X'y + b'X'Xb \\
&= \min_b y'y - 2y'Xb + b'X'Xb
\end{aligned}$$

where y is $(n \times 1)$ vector, X is $(n \times K)$ matrix whose rank is K and b is $(K \times 1)$ vector. Note that $y'X$ is a' vector, b is x vector and $X'X$ is A matrix in the formula above. Hence,

$$\begin{aligned}
\frac{S(b)}{\partial b} &= -2X'y + \left[(X'X) + (X'X)'\right]b \\
&= -2X'y + 2X'Xb
\end{aligned}$$

Therefore, we can find the solution to the minimization problem as;

$$b^* = (X'X)^{-1} X'y$$

You can check the second order condition for the minimum.

2. Some distribution theory

2.1 Multivariate normal distribution

- A random vector X with values in R^p is multivariate normal if every linear combination of its components $\alpha' X = \sum_{i=1}^{p} \alpha_i X_i$ follows a normal distribution on R.
- Every multivariate normal random vector has a finite mean $E(X) = \mu \in R^p$ and a finite covariance matrix $Var(X) = \Sigma$. We denote the random vector as $X \sim N(\mu; \Sigma)$.
- If $X \sim N(\mu; \Sigma)$; then $(AX + b) \sim N(A\mu + b; A\Sigma A')$ where A is an $(l \times p)$ matrix and b is an $(l \times 1)$ vector. Both of them are non-stochastic.
- If Σ is non-singular, then the density of X whose distribution is a multivariate normal is

$$f(x) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{\det \Sigma}} \exp\left[-\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right]$$

- Suppose we can partition X into X_1 and X_2 so that

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

Then , the marginal distributions of X_1 and X_2 are

$$X_1 \sim N(\mu_1; \Sigma_{11}) \text{ and } X_2 \sim N(\mu_2; \Sigma_{22})$$

- The conditional distribution are given by

$$X_1 \mid X_2 \sim N\left[\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2); \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right) \right]$$
$$X_2 \mid X_1 \sim N\left[\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1); \left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right) \right]$$

2.2 Censored and truncated normal distribution

- A censored normal random variable X is defined as

$$X = \begin{cases} X^* \text{ if } X^* > 0 \\ 0 \text{ otherwise} \end{cases}$$

where $X^* \sim N(\mu; \sigma^2)$. The censoring point is 0 here. But it can be any point on R.

- The density for X is given by

$$f(x) = \left[\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2 \right) \right]^{1-d} \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2 \right) \right]^{d}$$

$$= \left[\Phi\left(-\frac{\mu}{\sigma}\right)\right]^{1-d}\left[\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\right]^{d} = \left[1-\Phi\left(\frac{\mu}{\sigma}\right)\right]^{1-d}\left[\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\right]^{d}$$

where $d = 1$ when $X > 0$ and $d = 0$ when $X \le 0$. Φ is the cumulative distribution function of standard normal and ϕ is the density function of standard normal.

- Suppose that X is a censored normal random variable with censoring point of 0 and the distribution of the latent variable is given by $X^* \sim N(\mu, \sigma^2)$.

$$\text{(i)}\ E(X) = \sigma\phi\left(\frac{\mu}{\sigma}\right) + \mu\Phi\left(\frac{\mu}{\sigma}\right)$$

$$\text{(ii)}\ Var(X) = \sigma^2\left[\Phi\left(\frac{\mu}{\sigma}\right) + \frac{\mu}{\sigma}W\left(\frac{\mu}{\sigma}\right) - W^2\left(\frac{\mu}{\sigma}\right)\right]$$

where $W(z) = \int_{-1}^{z}\Phi(t)\,dt = z\Phi(z) + \phi(z)$

- A truncated normal random variable X is defined as

$$X = \begin{cases} X^* & \text{if } X^* > 0 \\ \text{not observed} & \text{otherwise} \end{cases}$$

where $X^* \sim N(\mu, \sigma^2)$. The truncation point is 0 here. But it can be any point on R.

- The density for X is given by

$$f(x) = \begin{cases} \dfrac{\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)}{\int_0^1 \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)} = \dfrac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{1-\Phi\left(-\frac{\mu}{\sigma}\right)} = \dfrac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} & \text{if } X > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Suppose that X is a truncated normal random variable with truncation point of 0 and the distribution of the latent variable is given by $X^* \sim N(\mu, \sigma^2)$.

$$\text{(i)}\ E(X) = \mu + \sigma\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}$$

$$\text{(ii)}\ Var(X) = \sigma^2\left[1 - \frac{\mu}{\sigma}\lambda\left(\frac{\mu}{\sigma}\right) - \lambda^2\left(\frac{\mu}{\sigma}\right)\right]$$

where $\lambda(z) = \frac{\phi(z)}{\Phi(z)}$ called the Mill's ratio.

2.3 Distribution derived from the normal distribution

2.3.1 χ^2 distribution

- χ^2 distribution with n degrees of freedom is the probability distribution of a random variable $Y = X_1 + X_2 + \cdots + X_n$, where the random variables X_i are independently distributed with respective distributions $N(\mu_i, 1)$. When $\mu_i = 0$ for all i, the chi-square distribution is said to be central, which usually called chi-square distribution without the adjective central and denoted as $\chi^2(n)$. It is non-central otherwise, which is called non-central chi-square distribution with non-centrality parameter $\lambda = \sum_{i=1}^{n}\mu_i^2$ and denoted as $\chi^2(n; \lambda)$.

- The density for a random variable X whose distribution is $\chi^2(n)$ is

$$f(x) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \exp\left(-\frac{x}{2}\right) 1_{(x>0)}$$

  note that $\chi^2(n)$ is a Gamma distribution with parameter $\frac{n}{2}$.
- When $X \sim \chi^2(n;\lambda)$, $E(X) = n+\lambda$ and $Var(X) = 2(n+2\lambda)$.
- Let Y be distributed as multivariate normal, $Y \sim N(\mu;\Sigma)$ where Σ is non-singular, and A be a symmetric matrix. Then, $Y'AY$ is distributed as a chi-square if and only if $A\Sigma A = A$, in which case the degrees of freedom is $\operatorname{rank}(A)$ and the non-centrality parameter is $\mu' A \mu$.
- For example, if X is an n-variate normal random variable, $X \sim N(\mu;\Sigma)$ with non-singular Σ, then $X'\Sigma^{-1}X \sim \chi^2\left(n;\mu'\Sigma^{-1}\mu\right)$. And, $(X-\mu)'\Sigma^{-1}(X-\mu) \sim \chi^2(n)$. We have chosen $A = \Sigma^{-1}$.

2.3.2 (Student) t distribution

- The t distribution with n degrees of freedom is the probability distribution of the random variable $Z = \frac{X}{\sqrt{\frac{Y}{n}}}$ where $X \sim N(\mu;1)$ and $Y \sim \chi^2(n)$. X and Y are independent each other. The parameter μ is called non-centrality parameter and the (central) t distribution corresponds to $\mu = 0$ and denoted as $t(n)$.
- The density of a random variable Z whose distribution is $t(n)$ is given by

$$f(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n}\,\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)\left(1+\frac{z^2}{n}\right)^{\frac{n+1}{2}}}$$

- If $Z \sim t(n)$, $E(Z^p) = 0$ when p is odd and $p < n$. $E(Z^p) = \frac{n^{\frac{p}{2}}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)}$ when p is even and $p < n$. When $p \geq n$, $E(Z^p)$ does not exist.
- Why do we need χ^2 and t? Here is a good example. Suppose X_i's are i.i.d. with the distribution $X_i \sim N\left(\mu;\sigma^2\right)$. It is well-known that $\frac{1}{n}\sum_{i=1}^n X_i = \overline{X} \sim N\left(\mu;\frac{\sigma^2}{n}\right)$. Moreover, $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$ where $s^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$. In addition, we can show that $\overline{X}$ and s^2 are independent. Hence,

$$\frac{\frac{\sqrt{n}}{\sigma}\left(\overline{X}-\mu\right)}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\frac{\sqrt{n}}{\sigma}\left(\overline{X}-\mu\right)}{\frac{s}{\sigma}} = \frac{\left(\overline{X}-\mu\right)}{\frac{s}{\sqrt{n}}} \sim t(n-1)$$

2.4 F distribution

- The F distribution with n_1 and n_2 degrees of freedom is the probability distribution of the random variable $W = \frac{Y_1/n_1}{Y_2/n_2}$ where $Y_1 \sim \chi^2(n_1)$ and $Y_2 \sim \chi^2(n_2)$. And, Y_1 and Y_2 are independent. The distribution is denoted as $F(n_1;n_2)$.

- The density for a random variable W whose distribution is $F(n_1, n_2)$ can be written as

$$f(w) = (n_1)^{\frac{n_1}{2}} (n_2)^{\frac{n_2}{2}} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right) w^{\frac{n_1}{2}-1}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)(n_2+n_1 w)^{\frac{n_1+n_2}{2}}}$$

- If $W \sim F(n_1, n_2)$, $E(W^p) = \left(\frac{n_2}{n_1}\right)^p \frac{\Gamma\left(\frac{n_1}{2}+p\right)\Gamma\left(\frac{n_2}{2}-p\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}$.
- We can define a non-central F distribution with two independent non-central chi-square distributions.
- If $Z \sim t(n)$, $Z^2 = \left(\frac{X}{\sqrt{\frac{Y}{n}}}\right)^2 = \frac{X^2/1}{Y/n} = F(1, n)$ since $X^2 \sim \chi^2(1)$, $Y \sim \chi^2(n)$ and X and Y are independent by definition.