Topic 2 : Background Material in Statistics

Dr. Ani Dasgupta
Department of International Business, Mass. Maritime Academy
and
Department of Economics, Boston University

1 Univariate Statistics Concepts You Must Know

This topic is mostly on multivariate statistics, and results relating to multivariate normal distribution (to be defined shortly). We marry matrix notation and statistical concepts to succintly express and prove many results used in both Ec507 and Ec508, which, as you will see, is quite neat. But before you proceed any further, you should stop and ask yourself whether you can answer each of the questions in the following list, for if you can't, instead of going further, you really should hit a good graduate (or advanced undergraduate) statistics text first.[1] Of particular importance, for the working satisticians are items 17-25. Do not worry if you do not all the answers right away, but do make it a point to eventually learn all of them, and quickly! Foundations are really, really important.

1. What are random variables and random vectors?
2. What do the terms 'joint distribution', 'marginal distribution' and 'conditional distribution' mean? If the joint pdf of 2 random variables is given as $f(x, y)$, can you derive the two marginal pdfs and the two conditional pdfs?
3. What does it mean to say that $X_1, \ldots, X_n$ are mutually independent random variables?
4. What are 'expectation', 'variance', 'conditional expectation' and 'conditional variance'?

[1]Here are three of my favorites: "Statistical Inference" by George Casella and Roger Berger, "Statistical Inference" by V. K. Rohatgi and "Introduction to Probability and Mathematical Statistics" by Lee Bain and Max Engelhardt. The first book is the market leader for graduate statistics text, the second one is an advanced undergraduate text with tons of examples, and the third one is somewhere in-between the first two in its level of rigor.

5. What is 'covariance' between two random variables? Can you show that independence of a pair of random variables implies zero covariance between them while the reverse is not true?

6. What is the 'Law of Iterated Expectation' or the 'Tower Property'?

7. Can you write down the probability mass functions of the following discrete distributions and calculate their means and variances: Bernoulli(p), Binomial(n, p), Poisson (λ), Geometric (p), Negative Binomial (n, p)?

8. What is the 'univariate normal distribution' and what are its moments (upto the fourth)?

9. What are the pdfs of Exponential (λ) Gamma (α, β) and Cauchy (m, b) distributions? What are their first and second moments?

10. How are the following random variables generated starting from a bunch of independent $N(0, 1)$ (z) variables: χ^2, t, F? In the latter two cases what are their 'degrees of freedom'?

11. What does the Weak Law of Large Numbers say?

12. What does the 'Central Limit Theorem' say? Can you provide an example of a population distribution for which the sample mean does not behave (approximately) normally even when the sample size is arbitrarily large?

13. What is 'sample mean' and why is it 'unbiased' for 'population mean'?

14. What is 'sample variance' and why is it 'unbiased' for 'population variance'?

15. What is a statistic and how does one derive the 'sampling distribution' of a statistic (say the sample mean) both theoretically and through computer simulations?

16. What do the following terms mean in the context of estimation : 'unbiasedness', 'efficiency', 'mean squared error' and 'consistency'?

17. What exactly is a 95% confidence interval for a parameter? Do you know how to calculate these confidence intervals for means and standard deviations of normal populations?

18. In the context of testing hypotheses, what are type I and type II errors and what is a critical region?

19. What are typical 'one-tailed' and 'two-tailed' hypotheses that are tested?

20. What is size of a test?

21. What is the power (function) of a test?

22. How do we test one and two-tailed hypotheses regarding the mean of a normally distributed population, with or without knowing the variance? What is the theory behind it?

23. How do we test the equality of variance of two normal populations? What is the theory behind it?

24. Assuming equality of variance between two normal population, how do we test equality of the two population means? What is the theory behind it?

25. Do you know how to do the tests described in the last three questions if you have a laptop with STATA but no statistical tables?

2 Random Vectors, Means and Covariances

By a random vector $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$ we simply refer to a vector whose components $y_1, \ldots y_p$ are all random variables. Its mean $\boldsymbol{\mu}$ is defined as

$$\boldsymbol{\mu} = E(\mathbf{y}) = \begin{bmatrix} Ey_1 \\ \vdots \\ Ey_p \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} \tag{1}$$

Expectation of a random matrix $\mathbf{X}$, similarly is given by the matrix whose (i, j)th element is $E(X_{ij})$.

The variance-covariance matrix (also known as variance matrix/ covariance matrix/ dispersion matrix) is defined as

$$\mathbf{\Sigma} = V(\mathbf{y}) = Cov(\mathbf{y}) = \begin{bmatrix} E(y_1-\mu_1)^2 & E(y_1-\mu_1)(y_2-\mu_2) & \ldots & E(y_1-\mu_1)(y_p-\mu_p) \\ \vdots & \vdots & \ddots & \vdots \\ E(y_p-\mu_p)(y_1-\mu_1) & E(y_p-\mu_p)(y_2-\mu_2) & \ldots & E(y_p-\mu_p)^2 \end{bmatrix} \tag{2}$$

Thus the (i, j) th entry of $V(\mathbf{y})$ is $E(y_i - \mu_i)(y_j - \mu_j)$, or the covariance between y_i and y_j (thus the i-th diagonal entry is $var(y_i)$. In particular, note that if $\mathbf{y} = (y_1, \cdots, y_n)'$, where $y_1, \cdots, y_n$ are iid with $var(y_i) = \sigma^2$, then $V(\mathbf{y}) = \sigma^2\mathbf{I}$. This follows of course, because covariance between independent random variables is zero.

The expectation operator works as linearly with vectors and matrices as it does with scalars. Using that, you should be able to verify that

Fact 1 $\mathbf{\Sigma} = E\left[(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})'\right] = E(\mathbf{y}\mathbf{y}') - \boldsymbol{\mu}\boldsymbol{\mu}'$.

If $\mathbf{x}$ and $\mathbf{y}$ are two $p \times 1$ and $q \times 1$ random vectors, we define their covariance matrix $Cov(\mathbf{x}, \mathbf{y})$ or $\mathbf{\Sigma_{xy}}$ analogously, i.e. it is the matrix the (i, j)th element of which is $E[(x_i - Ex_i)(y_j - Ey_j)]$. You can check that another way of saying this is $\mathbf{\Sigma_{xy}} = E[(\mathbf{x} - \boldsymbol{\mu_x})(\mathbf{y} - \boldsymbol{\mu_y})']$ where $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are the means of $\mathbf{x}$ and $\mathbf{y}$ respectively.

Using the definitions the following properties of expectation of linear functions of random vectors should be obvious:

Fact 2 *If* $\mathbf{y}$ *is a* $p \times 1$ *random vector and* $\mathbf{a}$ *is a* $p \times 1$ *vector made of non-random scalars, then*

$$E(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{E}(\mathbf{y}) \tag{3}$$

Fact 3 *If* $\mathbf{y}$ *is a* $p \times 1$ *random vector and* A *is a* $q \times p$ *matrix made of non-random scalars, then*

$$E(\mathbf{A}\mathbf{y}) = \mathbf{A}\mathbf{E}(\mathbf{y}) \tag{4}$$

Fact 4 *If* $\mathbf{X}$ *is a* $p \times q$ *random matrix and* $\mathbf{A}$, $\mathbf{B}$ *are non-random matrices such that* $\mathbf{AXB}$ *is well-defined, then*

$$E(\mathbf{AXB}) = \mathbf{AE}(\mathbf{X})\mathbf{B} \tag{5}$$

We now turn to formulas for calculating variance/covariance matrices of linear functions of random vectors. The notation is exactly as before, so I will not repeat it.

Fact 5

$$V(\mathbf{a}'\mathbf{y}) = \mathbf{a}'\mathbf{\Sigma}\mathbf{a} \quad (\textit{where}\, \mathbf{\Sigma} = V(\mathbf{y})) \tag{6}$$

Fact 6

$$V(\mathbf{A}\mathbf{y}) = \mathbf{A}\mathbf{\Sigma}\mathbf{A}' \quad (\textit{where}\, \mathbf{\Sigma} = V(\mathbf{y})) \tag{7}$$

Fact 7

$$Cov(\mathbf{a}'\mathbf{y}\, , \mathbf{b}'\mathbf{y}) = \mathbf{a}'\mathbf{\Sigma}\mathbf{b} \quad (\textit{where}\, \mathbf{\Sigma} = V(\mathbf{y})) \tag{8}$$

Fact 8

$$Cov(\mathbf{A}\mathbf{y}\, , \mathbf{B}\mathbf{y}) = \mathbf{A}\mathbf{\Sigma}\mathbf{B}' \quad (\textit{where}\, \mathbf{\Sigma} = V(\mathbf{y})) \tag{9}$$

Fact 9

$$Cov(\mathbf{A}\mathbf{x}\, , \mathbf{B}\mathbf{y}) = \mathbf{A}\mathbf{\Sigma}_{\mathbf{xy}}\mathbf{B}' \quad (\textit{where}\, \mathbf{\Sigma}_{\mathbf{xy}} = Cov(\mathbf{x},\, \mathbf{y})) \tag{10}$$

The above formulas (Facts 5 through 9) can be verified by writing out the (i, j)th element of both the left hand sides and the right hand sides.

Section End Questions

1. If $\mathbf{V}$ is the variance covariance matrix of the random vector $\mathbf{x}$ and $\mathbf{a}$ is some non-random vector, then the quadratic form $\mathbf{a}'\mathbf{V}\mathbf{a}$ is the variance of something. What is that something? Hence, argue that variance covariance matrices must be positive semidefinite. If perchance it is not positive definite, what must be going on?

2. Use the matrix formulas above to explicitly write down the formula for the variance of $a_1x_1 + a_2x_2 + \ldots + a_nx_n$ in terms of the a's, the variances of the x's and their (pairwise) covariances.

3. Use the matrix formulas above to explicitly write down the covariance between $ax + by$ and $cx + dy$ in yterms of a, b, c, d, the variances of x and y and their covariance.

3 Multivariate Normal Distribution

3.1 Definition

Multivariate normal distribution theory is important to us since in the classical linear model, the vector of random disturbance terms is often assumed to follow this distribution. This in turn characterizes the properties of the OLS estimates and various test statistics. Only because of this assumption, we can conduct the various t and F tests that we use to test hypotheses of interest. The exact form of the density function is needed again when we deal with maximum likelihood theory, which is used to establish certain desirable properties of the classical estimators under this distributional assumption.

The formula for the density function of a p-dimensional multivariate normal random vector $\mathbf{y}$ is given by

$$\frac{1}{(\sqrt{2\pi})^p|\boldsymbol{\Sigma}|^{1/2}}\, e^{-\frac{1}{2}\,(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})} \tag{11}$$

which takes two parameters: $\boldsymbol{\mu}$ which is a $p \times 1$ vector and $\boldsymbol{\Sigma}$ which is a $p \times p$ symmetric positive definite matrix. We will write this compactly as: $\mathbf{y} \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Although the equation above looks like a formidable expression, you can see that for the special case of $p = 1$ (when both μ and $\boldsymbol{\Sigma}$ are scalars), with a relabeling of $\boldsymbol{\Sigma}$ as σ^2, the formula reduces to standard univariate normal pdf: $\frac{1}{(\sqrt{2\pi}\sigma)}\, e^{-\frac{1}{2}\,(\frac{x-\mu}{\sigma})^2}$. The figure below illustrates the pdf of a bivariate normal vector:

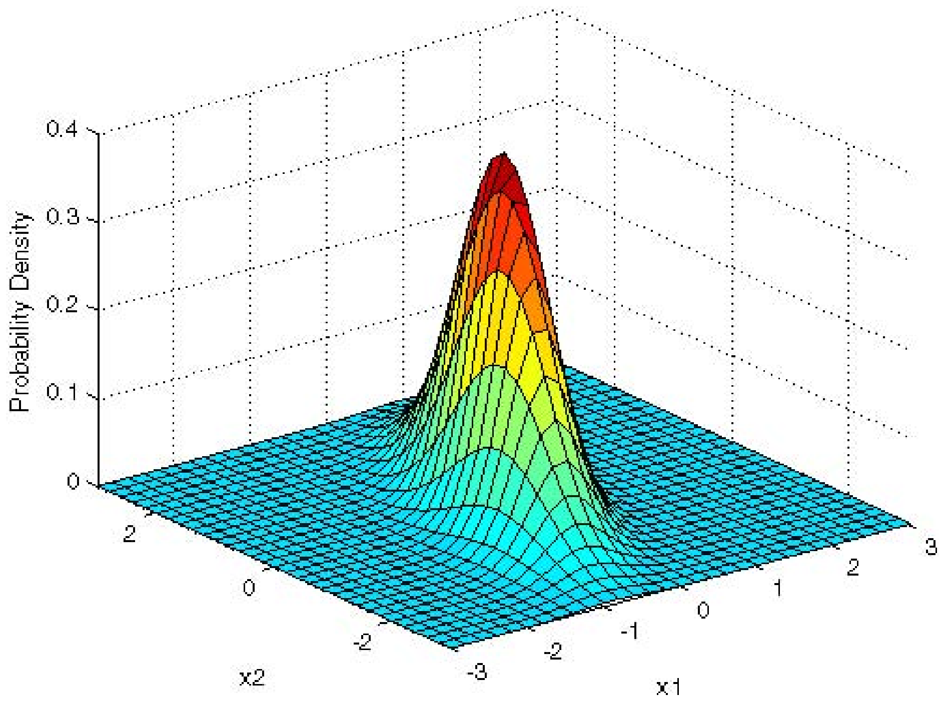

Figure 1. Multinormal Density with $\boldsymbol{\mu} = (0,0), \boldsymbol{\Sigma} = \begin{bmatrix} .25 & .3 \\ .3 & 1 \end{bmatrix}$ (via MATLAB)

Drawing analogy with the univariate case if you are thinking that $\boldsymbol{\mu}$ must be the mean vector of $\mathbf{y}$ and $\boldsymbol{\Sigma}$ must be its variance covariance matrix, you are exactly right. Let us prove this result in a slightly circuitous way, which will have the advantage of showing how we could generate a multivariate normal vector using batches of independent normal random variables.

Let $z_1, z_2, \ldots z_p$ be i.i.d. $N(0,1)$ random variables. Since they are independent, their joint density $g(z_1, z_2, \ldots z_p)$ is simply the product of the individual density functions.

$$\begin{aligned} g(z_1, z_2, \cdots z_p) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}} \cdots \frac{1}{\sqrt{2\pi}} e^{-\frac{z_p^2}{2}} \\ &= \frac{1}{(\sqrt{2\pi})^p} e^{\left(-\frac{z_1^2 + \ldots + z_p^2}{2}\right)} \\ &= \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{\mathbf{z}'\mathbf{z}}{2}} \qquad (12) \end{aligned}$$

where $\mathbf{z} = (z_1, \ldots z_p)'$. Notice that we could have written $\mathbf{z} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$ where $\mathbf{I}$ is the $p \times p$ identity matrix.

Now define the random vector $\mathbf{y} = \boldsymbol{\Sigma}^{\mathbf{1/2}}\mathbf{z} + \boldsymbol{\mu}$. Note that since $\boldsymbol{\Sigma}$ is the symmetric positive definite matrix, its square root matrix exists. Either type of square root discussed in Topic 1 will do here.

Then indeed $\mathbf{y}$ has mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. This is because $E(\mathbf{y}) = \boldsymbol{\Sigma}^{1/2}\, E(\mathbf{z}) + \boldsymbol{\mu} = \mathbf{0} + \boldsymbol{\mu} = \boldsymbol{\mu}$, and $Cov(\mathbf{y}) = \boldsymbol{\Sigma}^{1/2}\mathbf{I}(\boldsymbol{\Sigma}^{1/2})' = \boldsymbol{\Sigma}$ (recall that if $\mathbf{P}$ is the first type of square root, $\mathbf{PP}' = \mathbf{\Sigma}$ and if $\mathbf{Q}$ is the second type of square root, $\mathbf{Q}$ is symmetric and $\mathbf{QQ} = \mathbf{\Sigma}$). The rest involves showing that that the pdf of $\mathbf{y}$ is indeed given by equation (11).

To this end let us now recall the transformation formula you might be knowing from Ec 507; this formula allows us to find the (joint) density function of a set of random variables expressible in terms of another set of random variables the joint density function of the latter being known.

Transformation Formula: Let $y_1(z_1,\ldots,z_n)$, $y_2(z_1,\ldots,z_n),\ldots y_n(z_1,\ldots,z_n)$ define n new random variables in terms of n old random variables $z_1,\ldots,z_n$, so that the inverse transformations $z_1(y_1,\ldots,y_n)$, $z_2(y_1,\ldots,y_n),\ldots z_n(y_1,\ldots,y_n)$ are well-defined. Let $g(z_1,\ldots,z_n)$ denote the joint density function of $z_1,\ldots,z_n$. Then the joint density function of $(y_1,\ldots y_n)$ is given by (under mild regularity conditions)

$$f(y_1,\ldots,y_n) = Abs(|\mathbf{J}|)\, g(z_1(y_1,\ldots,y_n),\ldots\, z_n(y_1,\ldots,y_n)) \tag{13}$$

where the Jacobian matrix of the (*inverse*) transformation $\mathbf{J}$ is given by

$$\mathbf{J} = \begin{bmatrix} \frac{\partial z_1}{\partial y_1} & \cdots & \frac{\partial z_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_n}{\partial y_1} & \cdots & \frac{\partial z_n}{\partial y_n} \end{bmatrix}$$

and $Abs(|\mathbf{J}|)$ refers to the absolute value of its determinant.

Here is an example to illustrate how the formula works.

Example: Let z_1, z_2, z_3 be i.i.d. random variables with a common exponential density function $\phi(z) = e^{-z}$ (with $z > 0$). Thus $g(z_1, z_2, z_3) = e^{-(z_1+z_2+z_3)}$. Let

$$\begin{aligned} y_1 &= z_1 + z_2 + z_3 \\ y_2 &= \frac{z_1 + z_2}{z_1 + z_2 + z_3} \\ y_3 &= \frac{z_1}{z_1 + z_2} \end{aligned}$$

We are interested in figuring out the joint density function of (y_1, y_2, y_3).

The first step is to define the inverse transformations which are:

$$\begin{aligned} z_1 &= y_1 y_2 y_3 \\ z_2 &= y_1 y_2 - y_1 y_2 y_3 \\ z_3 &= y_1 - y_1 y_2 \end{aligned}$$

Hence,

$$\mathbf{J} = \begin{bmatrix} y_2 y_3 & y_1 y_3 & y_1 y_2 \\ y_2(1 - y_3) & y_1(1 - y_3) & -y_1 y_2 \\ 1 - y_2 & -y_1 & 0 \end{bmatrix}$$

If you calculate it, $|\mathbf{J}|$, the determinant turns out to be $-{y_1}^2 y_2$. Hence, $f(y_1, y_2, y_3) = {y_1}^2 y_2\, e^{-y_1}$ (since $e^{-z_1} e^{-z_2} e^{-z_3} = e^{-y_1}$). ♣

Now let us get back to the problem of determining the density of $\mathbf{y}$ where $\mathbf{y} = \mathbf{\Sigma^{1/2}z} + \boldsymbol{\mu}$, and $\mathbf{z} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$. Using the transformation formula, and basic properties of determinants (fill in the details!), we can see that

$$\begin{aligned} f(\mathbf{y}) &= g(\mathbf{z})|\mathbf{\Sigma}^{-1/2}| \\ &= g(\mathbf{z})|\mathbf{\Sigma}|^{-1/2} \\ &= \frac{1}{(\sqrt{2\pi})^p|\mathbf{\Sigma}|^{1/2}}\, e^{-\frac{\mathbf{z}'\mathbf{z}}{2}} \\ &= \frac{1}{(\sqrt{2\pi})^p|\mathbf{\Sigma}|^{1/2}}\, e^{-\frac{(\mathbf{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))'(\mathbf{\Sigma^{-1/2}}(\mathbf{y}-\boldsymbol{\mu}))}{2}} \\ &= \frac{1}{(\sqrt{2\pi})^p|\mathbf{\Sigma}|^{1/2}}\, e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\mathbf{\Sigma^{-1}}(\mathbf{y}-\boldsymbol{\mu})} \qquad (14) \end{aligned}$$

which is exactly what we needed to show.

A slightly more general definition of the multivariate normal distribution can be given in terms of characteristic functions; this will be done in the last section of this chapter.

3.2 Properties

We now state a few important facts about the multivariate normal distribution.

Fact 10 *Linear Functions of normally distributed random vectors are normally distributed (with means and variances given by the formulas in the previous section). Hence, if* $\mathbf{x} \sim \mathbf{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$*, and* $\mathbf{A}$ *is any* $q \times p$ *matrix with full row rank* q *with* $q \leq p$*, then* $\mathbf{Ax} \sim \mathbf{N}(\mathbf{A}\boldsymbol{\mu}, \mathbf{A\Sigma A'})$*.*[2]

In what follows, suppose $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathbf{N}\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} , \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix} \right]$

Fact 11 *In the Fact above,* $\mathbf{x_1}, \mathbf{x_2}$ *are themselves normally distributed with* $\mathbf{x}_i \sim \mathbf{N}(\boldsymbol{\mu}_i, \mathbf{\Sigma}_{ii})$*. Thus normal vectors have normal marginals.*

Fact 12 *If* $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21} = \mathbf{0}$*, then* $\mathbf{x_1}$ *and* $\mathbf{x_2}$ *are independent. Thus zero covariance/correlation among subvectors implies their independence for normally distributed vectors.*

Fact 13 *Normal vetors have normal conditionals (conditioned on other components of the vector). Specifically,* $\mathbf{x_2}|\mathbf{x_1}$ *is normal with the conditional expectation and variance*

[2] The Fact is actually true for any arbitrary q, and $\mathbf{A}$ does not need to have full row rank. The reason these qualifiers are used here is that otherwise $\mathbf{A\Sigma A'}$ is not necessarily positive definite which was a required property of the variance covariance matrix of a normal distribution as we have defined it so far. The definition of the normal distribution and hence this property can be generalized as we will soon see.

given by the following formulas:

$$E[\mathbf{x_2}|\mathbf{x_1}] = E(\mathbf{x_2}) + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}(\mathbf{x_1} - E\mathbf{x_1}) \tag{15}$$

$$Cov(\mathbf{x_2}|\mathbf{x_1}) = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}{}^{-1}\mathbf{\Sigma}_{12} \tag{16}$$

Section End Questions

1. Show that a normal random vector translated by a (non-random) vector stays normal.
2. If $(x, y)'$ is jointly normal is x normal, given $x + y = 10$?
3. x and y are jointly normal random variables with variance 1 each and covariance -.5. Create two normal random variables u and v both of which are linear combinations of x and y such that they are independent.

4 Distribution of Quadratic Forms

4.1 Theory

We now state and prove several important theorems about the distributions of objects which look like $\mathbf{z'Az}$ where $\mathbf{z}$ is a multivariate normal $n-$vector and $\mathbf{A}$ is some $n \times n$ matrix. These results constitute the heart of inference in linear models including all the t-tests and F-tests you have encountered or are likely to encounter.

Theorem 1 *Let* $\mathbf{x}$ *be a random n-vector with* $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$*. Then* $\mathbf{x'x} \sim \chi^2_n$.

Proof: $\mathbf{x'x} = x_1^2 + x_2^2 + \ldots x_n^2$; each $x_i \sim N(0, 1)$, and x_i , x_j are independent for $i \neq j$. Hence the theorem follows from the definition of the Chi-squared distribution. ♣

Theorem 2 *If* $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$*, then* $\mathbf{x'\Sigma^{-1}x} \sim \chi^2_n$.

Proof: Since $\mathbf{\Sigma}$ is symmetric and positive definite, it admits a factorization, $\mathbf{\Sigma} = \mathbf{PP'}$. Note that this implies: $\mathbf{P}^{-1}\mathbf{\Sigma}\,(\mathbf{P})^{-1'} = \mathbf{P}^{-1}\mathbf{\Sigma}\,(\mathbf{P'})^{-1} = \mathbf{P}^{-1}\mathbf{PP'}(\mathbf{P'})^{-1} = \mathbf{I}$.

Now let $\mathbf{y} = \mathbf{P^{-1}x}$. Clearly, then $\mathbf{y}$ is multivariate normal with $E(\mathbf{y}) = \mathbf{0}$. Also, $V(\mathbf{y}) = \mathbf{P}^{-1}\,V(\mathbf{x})(\mathbf{P}^{-1})' = \mathbf{P}^{-1}\mathbf{\Sigma}\,(\mathbf{P}^{-1})' = \mathbf{I}$. Hence, by Theorem 1, $\mathbf{y'y} \sim \chi^2_n$. But $\mathbf{y'y} = \mathbf{x'(P^{-1})'P^{-1}x} = \mathbf{x'(P')^{-1}P^{-1}x} = \mathbf{x'(PP')^{-1}x} = \mathbf{x'\Sigma^{-1}x}$. Hence, $\mathbf{x'\Sigma^{-1}x} \sim \chi^2_n$. ♣

It follows that if $\mathbf{x} \sim \mathbf{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, then $(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma^{-1}}(\mathbf{x} - \boldsymbol{\mu}) \sim \chi^2_n$.

Theorem 3 *If the random n-vector $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$, and $\mathbf{A}$ is a symmetric $n \times n$ idempotent matrix of rank r, then $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2_r$.*

Proof: Let $\mathbf{Q}$ represent the matrix of eigenvectors of $\mathbf{A}$. Then $\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{\Lambda}$, the diagonal matrix of eigenvalues of $\mathbf{A}$. Since, $\mathbf{A}$ is idempotent, and is of rank r, we know that $\mathbf{\Lambda}$ has exactly r '1's and $n - r$ '0's along its diagonal.

Now let $\mathbf{y} = \mathbf{Q}'\mathbf{x}$. Clearly $E\mathbf{y} = \mathbf{0}$ and $V(\mathbf{y}) = \mathbf{Q}'\mathbf{I}\mathbf{Q} = \mathbf{I}$; hence, $y_1, \ldots y_n$ are independent, standard normal variates. Also, $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\mathbf{Q}'\mathbf{A}\mathbf{Q}\mathbf{y} = \mathbf{y}'\mathbf{\Lambda}\mathbf{y}$, an expression that contains sum of exactly r of the y_i^2's (the others don't show up because of the zero-s in $\mathbf{\Lambda}$), which implies that it is distributed as χ^2 variate with r degrees of freedom. ♣

More generally, if $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and $\mathbf{A}$ is idempotent with rank r, then $\frac{1}{\sigma^2}\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2_r$. Even more generally, if $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{A}$ is of rank r, and $\mathbf{A}\mathbf{\Sigma}$ is idempotent, then $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2_r$. You will be asked to prove this in a problem set problem.

Now we state two theorems about independence between functions of random vectors.

Theorem 4 *If $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$, and $\mathbf{A}$ and $\mathbf{B}$ are symmetric and idempotent matrices, then $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{x}'\mathbf{B}\mathbf{x}$ are independent if $\mathbf{A}\mathbf{B} = \mathbf{0}$.*

Proof: First note that using symmetry and idempotency of $\mathbf{A}$ and $\mathbf{B}$, we can write: $\mathbf{x}'\mathbf{A}\mathbf{x} = (\mathbf{x}'\mathbf{A}')(\mathbf{A}\mathbf{x})$ and $\mathbf{x}'\mathbf{B}\mathbf{x} = (\mathbf{x}'\mathbf{B}')(\mathbf{B}\mathbf{x})$. Clearly, therefore $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{x}'\mathbf{B}\mathbf{x}$ will be independent, if $\mathbf{A}\mathbf{x}$ and $\mathbf{B}\mathbf{x}$ are. This is indeed true as is shown in section 5 .... ♣

Theorem 5 *If $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$, and $\mathbf{A}$ is a symmetric and idempotent matrices, and $\mathbf{L}$ is any other matrix (such that $\mathbf{L}\mathbf{x}$ is defined) then $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{L}\mathbf{x}$ are independent if $\mathbf{L}\mathbf{A} = \mathbf{0}$.*

Sketch of Proof: Follows the same idea as of the previous proof by arguing that $\mathbf{L}\mathbf{x}$ and $\mathbf{A}\mathbf{x}$ are independent.... ♣

The last theorem I wish to state in this context is the famous Fisher-Cochran Theorem which deals with the interrelations among several quadratic forms. For a full proof of the general case I refer you to C.R. Rao's 1973 book, "Linear Statistical Inference and Its Applications".

Theorem 6 *Let $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$, and $\mathbf{A}_i$, $(i = 1, \ldots, k)$ are symmetric matrices. Let $\mathbf{A} = \sum_{i=1}^k \mathbf{A}_i$, rank$(\mathbf{A}_i) = r_i$ and rank $(\mathbf{A}) = r$. If any two of the following are true:*
(a) $\mathbf{A}_i$ is idempotent for all i
(b) $\mathbf{A}$ is idempotent
(c) $\mathbf{A}_i\mathbf{A}_j = 0$ for $i \neq j$,
then the following hold:
(1) $\frac{1}{\sigma^2}\mathbf{x}'\mathbf{A}_i\mathbf{x} \sim \chi^2_{r_i}$.
(2) $\mathbf{x}'\mathbf{A}_i\mathbf{x}$ and $\mathbf{x}'\mathbf{A}_j\mathbf{x}$ are independent.
(3) $\frac{1}{\sigma^2}\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^2_r$.

4.2 Application: 1 - sample t-test

Let $x_1, \ldots, x_n$ be i.i.d. normal variates with mean μ and variance σ^2. You must be well-aware that to test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$, we make use of the following t-statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $\bar{x}$, of course, is the sample mean and s is the sample standard deviation given by

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

To do our testing, we are told that the statistic t has a $t-$distribution with $n-1$ degrees of freedom. The question is how do we know that? And why $n-1$ degrees of freedom (rather than n)?

These questions are answered if we can show the following:
(i) $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$.
(ii) $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2} \sim \chi^2_{n-1}$.
(iii) The two random variables in i) and ii) are independent.

The rest then follows from the definition of a $t-$distribution.

Define

$$\mathbf{y} = \frac{1}{\sigma} \begin{bmatrix} x_1 - \mu_0 \\ \vdots \\ x_n - \mu_0 \end{bmatrix}$$

It should be clear that $\mathbf{y} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$.

Now (i) can be proven by noticing that $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}}\boldsymbol{\iota}'\mathbf{y}$. Computing the mean and the variance of this random variable in question using standard formulas and by appealing to fact 10 (to obtain its normality) one sees that (i) follows.

To prove (ii), let

$$\mathbf{D} = \mathbf{I} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}'$$

which of course, is our old friend the deviation-producing matrix (which is symmetric and idempotent). Now note that

$$\begin{aligned} \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2} &= \sum_{i=1}^{n}\left(\frac{x_i - \mu_0}{\sigma} - \frac{\bar{x} - \mu_0}{\sigma}\right)^2 \\ &= (\mathbf{D}\mathbf{y})'(\mathbf{D}\mathbf{y}) \\ &= \mathbf{y}'\mathbf{D}\mathbf{y} \end{aligned} \tag{17}$$

which of course is a quadratic form. Now, Theorem 3 applies, so we know that $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma^2}$ is χ^2 distributed with as many degrees of freedom as is the rank of $\mathbf{D}$. But, since, the rank of an idempotent matrix is just its trace, all we need to do is to sum the diagonal elements of $\mathbf{D}$ to get the d.f. of the relevant χ^2 distribution. Since each of those elements is $1 - 1/n$, the diagonal entries sum to $n - 1$ and we are done proving (ii).

To prove (iii), we recall that the numerator of our test statistic is $\frac{1}{\sqrt{n}}\boldsymbol{\iota}'\mathbf{y}$ and the chi-squared variate we are taking square root of in the denominator is $\mathbf{y}'\mathbf{D}\mathbf{y}$. But, $\iota'\mathbf{D} = 0$, as you should be able to verify. Now Theorem 5 applies and we are done proving (iii).

The rationale behind 2-sample t-tests can also be built from the preceding theory. You should try to verify this yourself. Also, in the classical regression model, we will use very similar kinds of arguments to derive our t and F tests.

Section End Questions

1. Prove the following special case of the Fisher-Cochran theorem. If $\mathbf{A_1}, \mathbf{A_2}$ are symmetric idempotent matrices with ranks r_1, r_2 respectively, and if $\mathbf{A_1}\mathbf{A_2} = \mathbf{A_2}\mathbf{A_1} = \mathbf{0}$, then $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a chi-squared variate with degree of freedom $r_1 + r_2$ where $\mathbf{A} = \mathbf{A_1} + \mathbf{A_2}$ and $\mathbf{x} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$.
2. Verify $\iota'\mathbf{D} = 0$ without a single line of algebra.
3. That sample mean and sample variance are independent is special to normal variates. Can you demonstrate some other population distribution for which this property does not hold? You need to explicitly demostrate the lack of independence of the two sample statistics.
4. Provide the rationale behind the 2-sample t-test.

5 Characteristic Functions and More On Multivariate Normal Distribution

This section provides a more general definition of the multivariate normal distribution that is often useful. The definition uses the notion of characteristic functions.

The characteristic function of a random vector $\mathbf{x}$ is a vector function $\phi(\mathbf{t}) = Ee^{(i\,\mathbf{t}'\mathbf{x})} = E(cos(\mathbf{t}'\mathbf{x})) + iE(sin(\mathbf{t}'\mathbf{x}))$. Characteristic functions always exist (in contrast to moment generating functions which are given by $Ee^{(\mathbf{t}'\mathbf{x})}$ and which typically get wider coverage in statistics texts) and it is a well-known result from probability theory that two random vectors have the same characteristic function if and only if they have the same cdf. Characteristic functions are easier to tackle than distribution functions when you are considering sums of independent random variables because $Ee^{(i\,\mathbf{t}'(\mathbf{x}+\mathbf{y}))} = Ee^{(i\,\mathbf{t}'\mathbf{x})}\,Ee^{(i\,\mathbf{t}'\mathbf{y})}$ when $\mathbf{x}, \mathbf{y}$ are independent.[3] Also, it provides a tool for checking whether random vectors are independent via the following result: $\mathbf{x}, \mathbf{y}$ are independent if for all $\mathbf{t}, \mathbf{q}$,

$$Ee^{(i\,(\mathbf{t}'\mathbf{x}+\mathbf{q}'\mathbf{y}))} = Ee^{(i\,\mathbf{t}'\mathbf{x})}\,Ee^{(i\,\mathbf{q}'\mathbf{y})} \tag{18}$$

[3] In contrast, finding the cdf of sums of random vectors even when they are independent is often quite challenging.

We will shortly put this result to use.

A more general definition of multivariate normal distribution is this: *A random vector is multivariate normal with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$, if its characteristic function computes to $\phi(\mathbf{t}) = e^{i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$.*

The reason this is a more general definition than the one given before is that here we allow for $\boldsymbol{\Sigma}$ to be positive semidefinite rather than positive definite.[4] This permits definition of a so-called singular normal distribution of random vectors where, for instance, it could be that all realizations of a 3-dimensional random vector are on a two dimensional plane (for example, consider $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)'$ where the $\mathbf{u}_1$ and $\mathbf{u}_2$ are independent normals and $\mathbf{u}_3 = \mathbf{u}_1 + \mathbf{u}_2$). In case $\boldsymbol{\Sigma}$ is positive definite, it can be shown that the pdf of a random vector with the aforementioned characteristic function is indeed given by the pdf described in section 3.1.

This definition permits another very elegant characterization of the multivariate normal distribution: *A random vector $\mathbf{x}$ is multivariate normal if and only if for all non-random vectors $\mathbf{l}$, $\mathbf{l}'\mathbf{x}$ is univariate normal.*

In the light of our new definition, Fact 10 may be stated more generally by omitting the restriction on $\mathbf{A}$; it need not be a full row-rank matrix. Also, given this fact and the independence-checking criterion provided by equation 18, we can complete the proof of Theorems 4 and 5 easily. For instance, to show $\mathbf{Ax}$ and $\mathbf{Bx}$ are independent in Theorem 4, we write

$$\begin{aligned} Ee^{(i\,(\mathbf{t}'\mathbf{Ax}+\mathbf{q}'\mathbf{Bx}))} &= Ee^{(i\,(\mathbf{t}'\mathbf{A}+\mathbf{q}'\mathbf{B})\mathbf{x})} \\ &= e^{i(\mathbf{t}'\mathbf{A}+\mathbf{q}'\mathbf{B}).\mathbf{0}-\frac{1}{2}(\mathbf{t}'\mathbf{A}+\mathbf{q}'\mathbf{B})\sigma^2\mathbf{I}(\mathbf{A}'\mathbf{t}+\mathbf{B}'\mathbf{q})} \\ &= e^{-\frac{1}{2}(\mathbf{t}'\sigma^2\mathbf{At})-\frac{1}{2}(\mathbf{q}'\sigma^2\mathbf{Bq})} \\ &= Ee^{i\,(\mathbf{t}'\mathbf{Ax})}Ee^{i\,(\mathbf{q}'\mathbf{Bx})} \end{aligned} \tag{19}$$

where we make use of the fact that $\mathbf{x}$ is normally distributed and also the form of the characteristic function of a normally distributed random vector in going from the first line to the second and the properties of $\mathbf{A}, \mathbf{B}$ in going from the second to the third. But now, the end product is easily seen to be the product of the characteristic functions of two normally distributed random vectors $\mathbf{Ax}$ and $\mathbf{Bx}$ and we are done.

Section End Questions

1. Prove the abovementioned ‘elegant characterization’ of multivariate normal random vectors.
2. Verify Fact 10 in its generality.

[4] If you did not answer the first section-end question of Section 2, you might wish to revisit it now.