Ragnar Frisch

ON THE DISTRIBUTION OF A SUM OF SQUARES OVER A LINEAR SUBSPACE

Sonderdruck aus
ZEITSCHRIFT FÜR ÖKONOMETRIE
1. Jahrgang, Juni 1950, Heft 1
Verlag der Gutenberg-Buchhandlung
Mainz · Universität

NOTIZEN

ON THE DISTRIBUTION OF A SUM OF SQUARES OVER A LINEAR SUBSPACE

By

RAGNAR FRISCH

Let the stochastic variables $x_1\, x_2 \ldots x_n$ be independently and normally distributed with expected values zero and with the same standard-deviation σ. As is well known the sum of squares

$$z = x_1^2 + x_2^2 + \ldots + x_n^2 \tag{1}$$

is distributed in the χ^2 distribution with n degrees of freedom. That is to say the density function (the probability density) of z is

$$f(z, n) = \frac{1}{2^{\frac{n}{2}} \sigma^n \Gamma\left(\frac{n}{2}\right)} z^{\frac{n}{2}-1} e^{-\frac{z}{2\sigma^2}} \quad (z > 0). \tag{2}$$

Consider any n—p-dimensional linear subspace L_{n-p} passing through origin, that is to say consider the locus of points whose coordinates $x_1\, x_2 \ldots x_n$ satisfy the p homogeneous linear equations

$$\begin{array}{l} a_{11}\, x_1 + a_{12}\, x_2 + \ldots + a_{1n}\, x_n = 0 \\ a_{21}\, x_1 + a_{22}\, x_2 + \ldots + a_{2n}\, x_n = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{p1}\, x_1 + a_{p2}\, x_2 + \ldots + a_{pn}\, x_n = 0 \end{array} \tag{3}$$

where the coefficients a form a non singular $p \cdot n$ matrix (i. e. a $p \cdot n$ matrix of rank p), but are otherwise arbitrary.

What is the density function (the probability density) of the sumsquare (1) over the subspace L_{n-p} defined by (3)? It is a fundamental fact that this distribution is simply the same χ^2 distribution as we had before, only with n—p degrees of freedom instead of n. And this holds good regardless of what the coefficients a actually are provided only that the rank of their matrix is p. In other words we have the following

(4) *Theorem: Let the n stochastic variables* $x_1\, x_2 \ldots x_n$ *be independently and normally distributed with expected values zero and all with the same standarddeviation* σ. *The density function (the probability density) of the sumsquare* (1) *over any (n—p)-dimen-*

sional linear subspace passing through origin (that is over any linear subspace of the form (3)*) is*

$$f(z,\nu) = \frac{1}{2^{\frac{\nu}{2}} \sigma^{\nu} \Gamma\left(\frac{\nu}{2}\right)} z^{\frac{\nu}{2}-1} e^{-\frac{z}{2\sigma^2}} \quad (z > 0) \tag{5}$$

where $\nu = n - p$.

In the statistical literature this theorem is always taken more or less for granted, but it is not easy to find a proof of it that proceeds step by step in a complete and clear cut fashion, it is even difficult to find a precise formulation of the theorem. It may therefore be worth while to go over this ground and give a chain of reasoning that leads up to the theorem[1] and exhibits in the simplest possible way the strategic points of the argument when (1) is assumed.

Obviously, since the x_i are *independent* we can immediately conclude that if we select any ν of them, say $x_\alpha, x_\beta, \dots x_\gamma$ where $\alpha, \beta, \dots \gamma$ is any set of ν numbers selected from the set $1, 2, \dots n$ $(1 \leqq \nu \leqq n)$, and we form the sumsquare

$$z = x_\alpha^2 + x_\beta^2 + \dots + x_\gamma^2, \tag{6}$$

then this z is distributed in the χ^2 distribution with ν degrees of freedom, that is, its density function (probability density) is given by putting $n = \nu$ in (2). Furthermore if we perform on the set of variables $x_1\, x_2, \dots x_n$ any orthogonal transformation so as to obtain a new set of variables $y_1\, y_2 \dots y_n$, it is a classical fact in the theory of the normal distribution that this new set will have the same property as the original set. That is to say all the y_i will be *independent*, have an expected value zero and all have the same standarddeviation, namely σ. We can therefore immediately conclude that if we select any ν of these new variables, say $y_\alpha, y_\beta, \dots y_k$, the sumsquare

$$z = y_\alpha^2 + y_\beta^2 + \dots + y_\gamma^2 \tag{7}$$

will also be distributed in the χ^2 distribution with ν degrees of freedom. That is to say the density function (probability density) of the stochastic variable z defined by (7) will be obtained simply by putting $n = \nu$ in (2). But these facts are not yet what is contained in theorem (4). The theorem is much broader. It tells us that if we change our attention from what happens in the total space, to what happens in *any* linear subspace through origin — which is the same as to limit our attention to what happens in that subunivers where the relations (3) are fullfilled — the only thing

[1] I acknowledge with thanks the critical as well as constructive remarks of Mr. Erling Sverdrup in the course of discussions on this problem.

we have to do in order to obtain the density function (probability density) of the stochastic variable z defined by (1) is to put n in (2) equal to the dimensionality of the subspace considered, that is equal to $n—p$ where p is the rank of the matrix of the coefficients a in (3).

As a preliminary to the proof consider a set of p vectors in n-dimensional space

$$\begin{aligned} a_1 &= (a_{11}\, a_{12} \ldots a_{1n}) \\ a_2 &= (a_{21}\, a_{22} \ldots a_{2n}) \\ &\ldots\ldots\ldots\ldots \\ a_p &= (a_{p1}\, a_{p2} \ldots a_{pn}) \end{aligned} \tag{8}$$

They are assumed to be linearly independent, i. e. to unfold a p-dimensional linear subspace L_p. This means that the $p \cdot n$ matrix (8) is assumed to be of rank p. The unfolding of the subspace L_p is achieved through the fact that any vector belonging to L_p can be written as a homogeneous linear form in the vectors (8).

For the subsequent applications it does not restrict generality if we assume all the vectors a to be normalized. That is to say we assume

$$a_{i1}^2 + a_{i2}^2 + \ldots + a_{in}^2 = 1 \quad ({}_i = 1, 2, \ldots p). \tag{9}$$

Apart from this and from the condition regarding the rank, the vectors a are entirely arbitrary.

The linear subspace *orthogonal* to L_p is the locus of points $(x = x_1\, x_2 \ldots x_n)$ such that (3) is fullfilled. In other words the two linear subspaces L_p and L_{n-p} are mutually orthogonal to each other.

Now consider the structure of the subspace L_p. There always exists a set of p *mutually orthogonal* vectors

$$\begin{aligned} c_1 &= (c_{11}\, c_{12} \ldots c_{1n}) \\ c_2 &= (c_{21}\, c_{22} \ldots c_{2n}) \\ &\ldots\ldots\ldots\ldots \\ c_p &= (c_{p1}\, c_{22} \ldots c_{pn}) \end{aligned} \tag{10}$$

that belong to L_p and unfold L_p. That these vectors are mutually orthogonal means that

$$c_{i1}\, c_{j1} + c_{i2}\, c_{j2} + \ldots + c_{in}\, c_{jn} = \begin{cases} (0 & i \neq j) \\ (1 & i = j) \end{cases} \tag{11}$$

$$\text{for } (i = 1, 2, \ldots p,\ j = 1, 2, \ldots p).$$

Since the vectors are orthogonal they must necessarily be linearly independent. The fact that they belong to L_p means that any of them can be expressed as a homogeneous linear form in the vectors (8). That they unfold L_p means that any vector that belong to L_p, i. e. which may be expressed as a homogeneous linear form in the p vectors (8) can also be expressed as a homogeneous linear form in the p vectors (10).

One way to construct such an orthogonal set of vectors is for instance the following. First put $c_1 = a_1$. Since a_1 is normalized, c_1 will be. Next put $c_2 = \alpha a_2 + \beta c_1$ and determine the constants α and β in such a way that c_1 and c_2 become orthogonal to each other and c_2 normalized. This leads to

$$\begin{gathered} \alpha \cdot (a_1 a_2) + \beta \cdot (a_1 a_1) = 0 \\ (\alpha a_2 + \beta a_1)^2 = 1 \end{gathered} \tag{12}$$

which gives

$$\begin{aligned} \alpha &= \frac{1}{\sqrt{(a_1 a_1)\ (a_2 a_2) - (a_1 a_2)^2}} \\ \beta &= \frac{-(a_1 a_2)}{\sqrt{(a_1 a_1)\ (a_2 a_2) - (a_1 a_2)^2}} \end{aligned} \tag{13}$$

Since the vectors a_1 and a_2 are linearly independent, the numerators in (10) (which are the Gramians of these two vectors) are positive, not zero. Thus the vector c_2 is uniquely determined, apart from the sign of the square root.

Next we put $c_3 = \gamma a_3 + \delta c_1 + \varkappa c_2$ and determine the three coefficients γ, δ, $\varkappa$ in such a way that c_3 becomes orthogonal both to c_1 and c_2 and also becomes a normalized vector. This leads to solvable equations in γ, δ, $\varkappa$. In this manner we may continue and determine all the vectors c.

Each vector c thus obtained can obviously be expressed as a homogeneous linear form in the vectors a and vice versa. Let these expressions be respectively

$$c_i = \lambda_{i1} a_1 + \lambda_{i2} a_2 + \ldots + \lambda_{ip} a_p \qquad (i = 1, 2, \ldots p) \tag{14}$$

$$a_i = \lambda_{i1}^{\chi} c_1 + \lambda_{i2}^{\chi} c_2 + \ldots + \lambda_{ip}^{\chi} c_p \qquad (i = 1, 2, \ldots p) \tag{15}$$

The fact that we have chosen the c-vectors in such a particular way that $\lambda_{ik} = \lambda_{ik}^{\chi} = 0$ for $k > i$ is not essential for the present argument, all that is needed is to show that *some* selection of the orthogonal c—s is possible.

This being so, the system of homogeneous equations (3) may also be written in another form, namely in the form

$$\begin{aligned} c_{11}\, x_1 + c_{12}\, x_2 + \ldots + c_{1n}\, x_n &= 0 \\ c_{21}\, x_1 + c_{22}\, x_2 + \ldots + c_{2n}\, x_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_{p1}\, x_1 + c_{p2}\, x_2 + \ldots + c_{pn}\, x_n &= 0 \end{aligned} \tag{16}$$

Indeed, if we multiply the first equation in (3) by λ_{i1}, the second by $\lambda_{i2} \ldots$ and the p—*th* by λ_{ip} and add up, the result will be the i—*th* equation in (16). This may be done for all $i = 1, 2, \ldots p$. Inversely, if we multiply the first equation in (16) by λ_{i1}^{χ}, the second by $\lambda_{i2}^{\chi} \ldots$ and the p—*th* by λ_{ip}^{χ}

and add up, the result will be the *i—th* equation in (3). And this too may be done for all $i = 1, 2, \ldots p$. This means that the two systems (3) and (16) are *equivalent* in the sense that any set of values of $x_1\, x_2 \ldots x_n$ which satisfy (3) must also satisfy (16) and vice versa. In a general way this fact can be interpreted by saying that it is the linear subspace L_p *as such* that is important, not the particular reference system that is used to unfold it. Some of these reference systems — the orthogonal ones — are for our purpose more convenient than the others. In any case the role of the subspace L_p is that it is the "normal" by means of which the subspace L_{n-p} is defined.

From now on we take (16) instead of (3) as the definition of the region L_{n-p} within which we want to study the distribution of the stochastic variable z defined by (1).

We introduce a set of n—p *further* vectors

$$\begin{aligned} c_{p+1} &= (c_{p+1,1}\, c_{p+1,2} \ldots c_{p+1,n}) \\ c_{p+2} &= (c_{p+2,1}\, c_{p+2,2} \ldots c_{p+2,n}) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ c_n &= (c_{n,1}\ \ c_{n,2}\ \ \ldots\ \ c_{n,n}) \end{aligned} \tag{17}$$

which are mutually orthogonal as well as orthogonal against any of the vectors in (10), but are otherwise arbitrary. That such a system of vectors exists can be seen in many ways, for instance by adding to (8) *any* set of n—p further vectors $a_{p+1}, a_{p+2}, \ldots a_n$ such that the complete set $a_1, a_2, \ldots a_n$ becomes of rank n (and, for convenience, such that (9) is fullfilled also for $i = p + 1, p + 2 \ldots n$) and then constructing the set $c_{p+1}, c_{p+2} \ldots c_n$ simply by continuing the same process as was used for the construction of $c_1\, c_2 \ldots c_p$. Incidentally, the arbitrariness of the vectors $a_{p+1}, a_{p+2} \ldots a_n$ exhibits the variety of ways in which the set $c_{p+1}, c_{p+2} \ldots c_n$ may be constructed.

Now consider the orthogonal transformation on the *complete* set of variables $x_1\, x_2 \ldots x_n$ which is defined through the n vectors $c_1\, c_2 \ldots c_n$ (linearly independent since they are orthogonal). That is to say consider the (non singular) transformation

$$\begin{aligned} y_1 &= c_{11}\, x_1 + c_{12}\, x_2 + \ldots + c_{1n}\, x_n \\ y_2 &= c_{21}\, x_1 + c_{22}\, x_2 + \ldots + c_{2n}\, x_n \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n &= c_{n1}\, x_1 + c_{n2}\, x_2 + \ldots + c_{nn}\, x_n \end{aligned} \tag{18}$$

Since the complete transformation is orthogonal all the variables $y_1\, y_2 \ldots y_n$ will be *independent* with expected value zero and standard deviation σ. Hence if we select any n—p of them, say $y_{p+1}, y_{p+2}, \ldots, y_n$, the sumsquare of these variables, that is

$$z = y_{p+1}^2 + y_{p+2}^2 + \ldots + y_n^2 \tag{19}$$

will be distributed in the χ^2 distribution with $n-p$ degrees of freedom. But, *within the region defined* by (3) the quantity z defined by (19) is *identical* with the quantity z defined by (1). Indeed for any values of $x_1\, x_2 \ldots x_n$ in total space we have

$$y_1^2 + y_2^2 + \ldots + y_n^2 = x_1^2 + x_2^2 + \ldots + x_n^2 \tag{20}$$

because

$$\Sigma_k y_k^2 = \Sigma_k \Sigma_{ij}\, c_{ki}\, x_i \cdot c_{kj} x_j = \Sigma_{ij} x_i x_j\, \Sigma_k e_{ki} c_{kj} = \Sigma_i\, x_i^2,$$

and everywhere in the region defined by (3) we have

$$y_1 = y_2 = \ldots = y_p = 0 \tag{21}$$

In other words, *within* the region considered two things happen: First, the quantity (19) is here distributed in the χ^2 distribution with $n-p$ degrees of freedom, and second, this quantity is here identical with the quantity (1) whose distribution we want to determine. This establishes the theorem (4).

Zusammenfassung

Die stochastischen Variablen $x_1, x_2, \ldots, x_n$ seien unabhängig und normal verteilt mit Erwartungswerten 0 und sämtlich mit dem mittleren Fehler σ. Dann gilt für die Dichtefunktion (Wahrscheinlichkeitsdichte) der Quadratsumme (1) über irgendeinem $(n-p)$-dimensionalen linearen Teilraum (durch den Nullpunkt), d. h. über irgendeinem linearen durch (3) definierten Teilraum vom Range p, die Darstellung (5) mit $\nu = n - p$. Für diesen in der Statistik oft benutzten, aber anscheinend noch nicht vollständig bewiesenen Satz wird ein Beweis gegeben und sein Gehalt herausgearbeitet.

Résumé

Les variables stochastiques $x_1, x_2, \ldots, x_n$ soient indépendantes et distribuées normales avec l'espérance mathématique 0 et toutes de l'erreur moyenne σ. Pour le fonction de densité de la somme quadratique (1) sur une variété linéaire de la dimension $n-p$ (par l'origine) arbitraire, c. à. d. sur une variété linéaire arbitraire du rang p, définie par (3), on a l'expression (5) avec $\nu = n - p$. Pour ce théorème, assez souvent appliqué dans le statistique mais non pas demontré complètement auparavent, est donné une démonstration et montré l'importance.