SOME THEOREMS ON QUADRATIC FORMS AND NORMAL VARIABLES

ZONGWU CAI

March 7, 2025

1. The Multivariate Normal Distribution

The $n \times 1$ vector of random variables, y, is said to be distributed as a multivariate normal with mean vector μ and variance covariance matrix Σ (denoted $y \sim N(\mu, \Sigma)$), if the density of y is given by

$$f(y; \mu, \Sigma) = \frac{\mathrm{e}^{-\frac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)}}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}}.$$

Consider the special case, where $n = 1$, $y = y_1$, $\mu = \mu_1$, and $\Sigma = \sigma^2$,

$$f(y_1; \mu_1, \sigma) = \frac{1}{(2\pi)^{\frac{1}{2}} (\sigma^2)^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(y_1-\mu_1)\left(\frac{1}{\sigma^2}\right)(y_1-\mu_1)} = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(y_1-\mu_1)^2}{\sigma^2}}$$

is just the normal density for a single random variable.

2. Theorems on Quadratic Forms in Normal Variables

Theorem 1. If $y \sim N(\mu_y, \Sigma_y)$, then, the moment generation function (mgf) $M(t)$ is given by

$$M_y(t) = E\left[e^{t^\top y}\right] = \exp\left(t^\top \mu + t^\top \Sigma t/2\right)$$

for any t, an $n \times 1$ vector. Then,

$$z = Ay \ \sim \ N\left(\mu_z = A\mu_y; \Sigma_z = A\Sigma_y A^\top\right),$$

where $A_{r\times n}$ is a matrix of constants. Furthermore, $z_1 = Ay$ and $z_2 = By$ are independent if and only if ($\iff$) $A\Sigma_y B^\top = 0$.

Proof: $E(z) = E(Ay) = AE(y) = A\mu_y$ and $\mathrm{Var}(z) = E\left[(z - E(z))(z - E(z))^\top\right] = E\left[(Ay - A\mu_y)(Ay - A\mu_y)^\top\right] = E\left[A(y - \mu_y)(y - \mu_y)^\top A^\top\right] = AE(y - \mu_y)(y - \mu_y)^\top A^\top =$

$A\Sigma_y A^\top$. Also, the mgf of z is $M_z(t) = E\left[e^{t^\top z}\right] = E\left[e^{(t^\top A)y}\right] = M_y(b)$, where $b = A^\top t$, a $r \times 1$ vector. Thus, z is normally distribution. Finally, define

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} y = D_1 y,$$

where D_1 is defined a clear manner, $z_1 = Ay$ and $z_2 = By$. Thus, $z \sim N(D_1\mu_y, \Sigma_1 = D_1\Sigma_y D_1^\top)$. Now,

$$\Sigma_1 = D_1\Sigma_y D_1^\top = \begin{pmatrix} A\Sigma_y A^\top & A\Sigma_y B^\top \\ B\Sigma_y A^\top & B\Sigma_y B^\top \end{pmatrix} = \begin{pmatrix} A\Sigma_y A^\top & 0 \\ 0 & B\Sigma_y B^\top \end{pmatrix} \iff A\Sigma_y B^\top = 0.$$

Therefore, z_1 and z_2 are independent $\iff A\Sigma_y B^\top = 0$. Then, Ay and By are independent $\iff A\Sigma_y B^\top = 0$.

Example: Let $Y_1, \ldots, Y_n$ denote a random sample drawn from $N(\mu, \sigma^2)$. Then,

$$Y = \begin{pmatrix} Y_1 \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} \sim N\left[\begin{pmatrix} \mu \\ \cdot \\ \cdot \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma^2 & \ldots & 0 \\ \cdot & \sigma^2 & \cdot \\ \cdot & & \\ 0 & & \sigma^2 \end{pmatrix}\right].$$

Now, Theorem 1 implies that

$$\bar{Y} = \frac{1}{n}Y_1 + \cdots + \frac{1}{n}Y_n = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right) Y = AY \sim N\left(\mu, \sigma^2/n\right),$$

since

$$\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mu \quad \text{and} \quad \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)\sigma^2 I \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Theorem 2. Let the $n \times 1$ vector $y \sim N(0, I)$. Then $y^\top y \sim \chi^2(n)$[1].

Proof: Consider that each y_i is an independent standard normal variable. Write out $y^\top y$ in summation notation as

$$y^\top y = \sum_{i=1}^{n} y_i^2,$$

which is the sum of squares of n standard normal variables. By applying the mgf approach, one can establish easily the theorem.

[1] The mgf of $\chi^2(n)$ is $M(t) = (1 - 2t)^{-n/2}$ for $t < 1/2$.

Theorem 3. If $y \sim N(0, \sigma^2 I)$ and M is a symmetric idempotent matrix of rank m, then, $\frac{y^\top M y}{\sigma^2} \sim \chi^2(\operatorname{tr} M = m)$. Also, if y is a $n \times 1$ random variable and $y \sim N(\mu, \Sigma)$, then, $(y-\mu)^\top \Sigma^{-1}(y-\mu) \sim \chi^2(n)$.

Proof: Since M is symmetric it can be diagonalized with an orthogonal matrix Q. This means that

$$Q^\top M Q = \Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Furthermore, since M is idempotent all these roots are either zero or one. Thus we can choose Q so that Λ will look like

$$Q^\top M Q = \Lambda = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}.$$

The dimension of the identity matrix will be equal to the rank of M, since the number of non-zero roots is the rank of the matrix. Since the sum of the roots is equal to the trace, the dimension is also equal to the trace of M. Now let $v = Q^\top y$. Compute the moments of $v = Q^\top y$ as follows. $E(v) = Q^\top E(y) = 0$ and $\operatorname{Var}(v) = Q^\top \sigma^2 I Q = \sigma^2 Q^\top Q = \sigma^2 I$, since Q is orthogonal. Then, $v \sim N(0, \sigma^2 I)$. Now, consider the distribution of $y^\top M y$ using the transformation v. Since Q is orthogonal, its inverse is equal to its transpose. This means that $y = \left(Q^\top\right)^{-1} v = Qv$. Now, write the quadratic form as follows

$$\frac{y^\top M y}{\sigma^2} = \frac{v^\top Q^\top M Q v}{\sigma^2} = \frac{1}{\sigma^2} v^\top \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} v = \frac{1}{\sigma^2} \sum_{i=1}^{\operatorname{tr} M} v_i^2 = \sum_{i=1}^{\operatorname{tr} M} \left(\frac{v_i}{\sigma}\right)^2.$$

This is the sum of squares of $(\operatorname{tr} M)$ standard normal variables and so is a χ^2 variable with $\operatorname{tr} M$ degrees of freedom. Let $w = (y-\mu)^\top \Sigma^{-1}(y-\mu) = z^\top z$, where $z = \Sigma^{-1/2}(y-\mu)$. By Theorem 1, z is distributed as $N(0, I)$, then, the proof is complete.

Corollary: If the $n \times 1$ vector $y \sim N(0, I)$ and the $n \times n$ matrix A is idempotent and of rank m. Then, $y^\top A y \sim \chi^2(m)$.

Theorem 4. If $y \sim N(0, \sigma^2 I)$, M is a symmetric idempotent matrix of order n, and L is a $k \times n$ matrix, then Ly and $y^\top M y$ are independently distributed if $LM = 0$.

Proof: Similar to the proof of Theorem 1, define

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} L \\ M \end{pmatrix} y = D_4 y,$$

where D_4 is defined a clear manner, $z_1 = Ly$ and $z_2 = My$. According to Theorem 1, $z \sim N(0, \Sigma_4 = \sigma^2 D_4 D_4^\top)$. Now,

$$\Sigma_4 = \sigma^2 D_4 D_4^\top = \sigma^2 \begin{pmatrix} LL^\top & LM^\top \\ ML^\top & MM^\top \end{pmatrix} = \sigma^2 \begin{pmatrix} LL^\top & 0 \\ 0 & MM^\top \end{pmatrix}$$

by the assumption that $LM = 0$. Therefore, z_1 and z_2 are independent. Since $y^\top My = z_2^\top z_2$, then, Ly and $y^\top My$ are independent.

Theorem 5. Let the $n \times 1$ vector $y \sim N(0, I)$, let A be an $n \times n$ idempotent matrix of rank m, let B be an $n \times n$ idempotent matrix of rank s, and suppose $BA = 0$. Then, $y^\top Ay \sim \chi^2(m)$, $y^\top By \sim \chi^2(s)$, and they are independently.

Proof: By Theorem 3 both quadratic forms are distributed as chi-square variables. We need only to demonstrate their independence. To this end, similar to Theorem 1, define

$$z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} y = D_5 y,$$

where D_5 is defined a clear manner, $z_1 = Ay$ and $z_2 = By$. According to Theorem 1, $z \sim N(0, \Sigma_5 = D_5 D_5^\top)$. Now,

$$\Sigma_5 = D_5 D_5^\top = \begin{pmatrix} AA^\top & AB^\top \\ BA^\top & BB^\top \end{pmatrix} = \begin{pmatrix} AA^\top & 0 \\ 0 & BB^\top \end{pmatrix}$$

by the assumption that $BA = 0$. Therefore, z_1 and z_2 are independent. Since $y^\top Ay = z_1^\top z_1$ and $y^\top By = z_2^\top z_2$, then, $y^\top Ay$ and $y^\top By$ are independent.

Theorem 6. If $y \sim N(\mu, \Omega)$ where Ω is positive definite, then, $q_1 = y^\top Ay$ and $q_2 = y^\top B\, y$ are independently distributed if and only if ($\Longleftrightarrow$) $A\Omega B = 0$.

Proof of Sufficiency: First, define $z = \Omega^{-1/2} y$. Then, $z \sim N(\Omega^{-1/2}\mu, I)$. Now, similar to Theorem 5,

$$t = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} y = \begin{pmatrix} A \\ B \end{pmatrix} \Omega^{1/2} z = D_6 z,$$

where D_6 is defined a clear manner, $t_1 = A\Omega^{1/2} z$ and $z_2 = B\Omega^{1/2} y$. According to Theorem 1, $t \sim N(D_6 \Omega^{-1/2}\mu, \Sigma_6 = D_6 D_6^\top)$. Now,

$$\Sigma_6 = D_6 D_6^\top = \begin{pmatrix} A\Omega A^\top & A\Omega B^\top \\ B\Omega A^\top & B\Omega B^\top \end{pmatrix} = \begin{pmatrix} A\Omega A^\top & 0 \\ 0 & B\Omega B^\top \end{pmatrix}$$

by the assumption that $A\Omega B = 0$. Therefore, t_1 and t_2 are independent. Since $y^\top Ay = t_1^\top t_1$ and $y^\top By = t_2^\top t_2$, then, $y^\top Ay$ and $y^\top By$ are independent. Note that the proof of necessity is difficult and has a long history; see, for instance, Driscoll and Grundberg (1986) and Discroll and Krasnicka (1995).

Example (Continued). Let s^2 denote the sample variance of $Y_1, \ldots, Y_n$ iid from $N(\mu, \sigma^2)$, given by

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

which can be expressed as $s^2 = Y^\top M_\ell Y/(n-1)$, where $M_\ell = I - \ell(\ell\ell^\top)^{-1}\ell^\top$ with $\ell^\top = (1, \ldots, 1)$. Also, $\bar{Y} = \ell^\top Y/n$. It is easy to see from the above theorems that $(n-1)s^2/\sigma^2 \sim \chi^2(1)$ and $M_\ell \ell = 0$, which implies that $\bar{Y}$ and s^2 are independent.

References

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton: Princeton University Press.

Driscoll, M. F. and W. R. Grundberg (1986). A history of the development of Craig's theorem. *American Statistician* **40**(1):65-69.

Driscoll, M. F. and B. Krasnicka (1995). An accessible proof of Craig's theorem in the general case. *American Statistician* **49**(1):59-61.

Goldberger, A. S. (1964). *Econometric Theory.* New York: Wiley.

Goldberger, A. S. (1991). *A Course in Econometrics.* Cambridge: Harvard University Press.

Hocking, R. R. (1985). *The Analysis of Linear Models. Monterey.* Brooks/Cole.

Hocking, R. R. (1996). *Methods and Applications of Linear Models.* New York: Wiley.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* 2nd edition. New York: Wiley.

Schott, J. R. (1997). *Matrix Analysis for Statistics.* New York: Wiley.