❘Elemente der Mathematik

Teaching the Independence of $\overline{X}$ and S^2

Wiebe R. Pestman

Wiebe R. Pestman has studied mathematics, physics, and astronomy at the University of Groningen where he also got his doctor's degree. His mathematical interests are in functional analysis, probability and statistics, operator algebras and harmonic analysis. Aside from mathematics he likes modern painting and music.

Introduction

In introductory courses in mathematical statistics there is always the very moment where students are confronted with the fact that, in case of a sample $X_1, \ldots, X_n$ from a normally distributed population, the statistics $\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ are independent. To the audience this will seem to be odd at first sight since the quantity $\overline{X}$ occurs explicitly in the defining equation of S^2. For this reason, omitting a rigorous college proof of this statement might cause a slight feeling of annoyance among certain students. In spite of this, many modern textbooks on statistics (even the excellent work [5]) more or less surrender in this respect.

The theory of multivariate normal distributions (see for example [1] or [4]) provides a natural framework for a proof, but in introductory courses it is not always advisable to present it in that way. Alternatives are to be found for example in [8], [9] and [10], where proofs are given relying on the theory of characteristic functions or moment generating functions. Here one is in fact applying part of the Fourier (Laplace) machinery of mathematical analysis, which is not always familiar to starting statisticians. In this paper a very elementary proof based on linear algebra is given. As in [3] and [6] a technique involving the notion of an orthogonal linear map is exploited to prove a more general result. The necessary linear algebra is usually taught in first year undergraduate courses in mathematics.

In einer Stichprobe aus einer Normalverteilung sind der Mittelwert $\overline{X}$ und die Varianz S^2 statistisch unabhängig. Die üblichen Beweise für diese für den Anfänger etwas überraschende Tatsache sind ziemlich involviert. Im vorliegenden Beitrag liefert Wiebe Pestman einen Beweis, der nur einfache Resultate der Linearen Algebra benötigt. *ust*

Independence of linear statistics of normal samples

Let $M = (M_1, \ldots, M_p)$ and $N = (N_1, \ldots, N_q)$ be vector valued random variables, assuming values in $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively. Denote by (M, N) the variable $(M_1, \ldots, M_p, N_1, \ldots, N_q)$ assuming its values in $\mathbb{R}^{p+q}$. We say that M and N are statistically independent if for every (say open) set $A \subset \mathbb{R}^p$ and $B \subset \mathbb{R}^q$ one has

$$\mathbb{P}\big((M, N) \in A \times B\big) = \mathbb{P}(M \in A)\, \mathbb{P}(N \in B).$$

If M and N are statistically independent and f and g are continuous functions on $\mathbb{R}^p$ and $\mathbb{R}^q$, then $f(M)$ and $g(N)$ also are statistically independent. In particular, for all i, j the components M_i and N_j are independent, if M and N are so. (It is frequently misunderstood (see for example [8], p. 350) that the converse of this statement is *not* true; see [2], [7].)

Now suppose that $X_1, \ldots, X_n$ are independent $N(0, 1)$ distributed variables. Writing $X = (X_1, \ldots, X_n)$ and denoting by $\langle \cdot, \cdot \rangle$ the standard inner product on $\mathbb{R}^n$, one has

$$\mathbb{P}(X \in A) = \int_A c\, e^{-\frac{1}{2}\langle x, x\rangle} dx,$$

where $A \subset \mathbb{R}^n$ and $c = (2\pi)^{-n/2}$. If $Q : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal linear transformation then it can easily be proved that $\mathbb{P}(X \in A) = \mathbb{P}(X \in QA)$. We shall refer to this property by saying that the probability distribution of X is *rotation invariant*. (It is interesting to note, that rotation invariance can only occur in cases where the X_i are all $N(0, \sigma^2)$ distributed; see [7].)

Let $\mathfrak{B}$ be the linear space consisting of all linear combinations of $X_1, \ldots, X_n$. An inner product $(\cdot, \cdot)$ on $\mathfrak{B}$ is defined by

$$(M, N) = \mathrm{cov}(M, N),$$

where $\mathrm{cov}(M, N)$ denotes the covariance between M and N. The Euclidean space $\mathfrak{B}$ will be referred to as the *space of linear statistics*. Note that $\{X_1, \ldots, X_n\}$ is an orthonormal basis in $\mathfrak{B}$.

Lemma: *If the linear statistics $Y_1, \ldots, Y_n$ form an orthonormal basis in $\mathfrak{B}$, then the vector valued random variables $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ are identically distributed.*

Proof: Let $\Phi : \mathfrak{B} \to \mathfrak{B}$ be the linear isometry which converts the orthonormal basis $\{X_1, \ldots, X_n\}$ into the orthonormal basis $\{Y_1, \ldots, Y_n\}$. Denoting the matrix of Φ with respect to basis $\{X_1, \ldots, X_n\}$ by $[\Phi]$, we can write

$$Y_j = \sum_{i=1}^{n} [\Phi]_{ij} X_i.$$

The matrix $[\Phi]$ being orthogonal, this implies that there exists an orthogonal linear transformation $Q : \mathbb{R}^n \to \mathbb{R}^n$ such that $QX = Y$. Taking into account that the probability distribution of X is rotation invariant one has for every open set $A \subset \mathbb{R}^n$

$$\mathbb{P}(Y \in A) = \mathbb{P}(QX \in A) = \mathbb{P}(X \in Q^{-1}A) = \mathbb{P}(X \in A).$$

This proves that X and Y are identically distributed.

Main theorem: *Let $X_1, \ldots, X_n$ be independent normally distributed variables. Suppose $M = (M_1, \ldots, M_p)$ and $N = (N_1, \ldots, N_q)$, where $M_i, N_j \in \mathfrak{V}$ for all i, j. Then M and N are statistically independent iff* $\operatorname{cov}(M_i, N_j) = 0$ *for every i, j.*

Proof: If M and N are statistically independent then so are the components M_i and N_j. Consequently one has $\operatorname{cov}(M_i, M_j) = 0$ for all i, j.

Next, we prove the converse of this statement in the special case where $X_1, \ldots, X_n$ are independent and $N(0, 1)$ distributed. Let $\mathfrak{M}$ and $\mathfrak{N}$ in $\mathfrak{V}$ be the linear span of $M_1, \ldots, M_p$ and $N_1, \ldots, N_q$ respectively. The assumption that $\operatorname{cov}(M_i, N_j) = 0$ for all i, j implies that $\mathfrak{M} \perp \mathfrak{N}$ in $\mathfrak{V}$. It follows that $\mathfrak{V}$ can be decomposed as $\mathfrak{V} = \mathfrak{M} \oplus \mathfrak{N} \oplus \mathfrak{R}$, where $\mathfrak{R} = (\mathfrak{M} \oplus \mathfrak{N})^{\perp}$. Choose orthonormal bases $\{E_1, \ldots, E_s\}$, $\{F_1, \ldots, F_t\}$ and $\{G_1, \ldots, G_u\}$ in $\mathfrak{M}$, $\mathfrak{N}$ and $\mathfrak{R}$ respectively. Set $E = (E_1, \ldots, E_s)$, $F = (F_1, \ldots, F_t)$ and $G = (G_1, \ldots, G_u)$. By the lemma the random variables $X = (X_1, \ldots, X_n)$ and (E, F, G) are identically distributed. From this it follows that $(X_1, \ldots, X_{s+t})$ and (E, F) also are identically distributed. Observing that $(X_1, \ldots, X_s)$ and $(X_{s+1}, \ldots, X_{s+t})$ are statistically independent, one has for $A \subset \mathbb{R}^s$ and $B \subset \mathbb{R}^t$

$$\mathbb{P}\big((E, F) \in A \times B\big) = \mathbb{P}\big((X_1, \ldots, X_{s+t}) \in A \times B\big)$$
$$= \mathbb{P}\big((X_1, \ldots, X_s) \in A\big)\, \mathbb{P}\big((X_{s+1}, \ldots, X_{s+t}) \in B\big).$$

It is easy to see that this implies $\mathbb{P}\big((E, F) \in A \times B\big) = \mathbb{P}(E \in A)\, \mathbb{P}(F \in B)$, thus proving the statistical independence of E and F. The variable M (N) can be obtained from E (F) by linear transformation, so M and N are also statistically independent. In the general case where $X_1, \ldots, X_n$ are independent and X_i is, say $N(\mu_i, \sigma_i^2)$ distributed, one can pass to variables $\widetilde{X}_i = (X_i - \mu_i)/\sigma_i$ and reduce to the preceding case.

Remark: We have stated the main theorem in terms of two random variables M and N, having components in $\mathfrak{V}$. The theorem can easily be generalised to the case of an arbitrary number of vector valued variables, having components in $\mathfrak{V}$. A generalisation of this kind can be applied for example in the theory of normal analysis of variance whenever independence of linear statistics has to be proved. The premise in the theorem is equivalent to the requirement that the vectorial variable (M, N) enjoys a multivariate normal distribution. Details of all this can be found in [7].

Theorem A: *If $X_1, \ldots, X_n$ form a sample from a normally distributed population, then $\overline{X}$ and S^2 are statistically independent.*

Proof: Observe that $\overline{X}$ and $X_i - \overline{X}$ are in $\mathfrak{V}$ and that $\operatorname{cov}(\overline{X}, X_i - \overline{X}) = 0$. By the main theorem we conclude that $\overline{X}$ and $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$ are independent; it thus follows that $\overline{X}$ and S^2 are independent.

Next, let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from a $N(\mu_X, \sigma^2)$ and a $N(\mu_Y, \sigma^2)$ distributed population, respectively. Set

$$\overline{X} = \frac{1}{m} \sum_i X_i, \qquad\qquad \overline{Y} = \frac{1}{n} \sum_j Y_j,$$

$$S_X^2 = \frac{1}{m-1} \sum_i (X_i - \overline{X})^2, \quad S_Y^2 = \frac{1}{n-1} \sum_j (Y_j - \overline{Y})^2.$$

The pooled variance (S_p^2) of both samples together is understood to be

$$S_p^2 = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}.$$

Theorem B: *The variables $(\overline{X}, \overline{Y})$ and S_p^2 are statistically independent.*

Proof: Applying the main theorem, the independence of the vectors $(\overline{X}, \overline{Y})$ and $(X_1 - \overline{X}, \ldots, X_m - \overline{X}, Y_1 - \overline{Y}, \ldots, Y_n - \overline{Y})$ is easily verified. As in the proof of theorem A the independence of $(\overline{X}, \overline{Y})$ and S_p^2 follows.

Remark: Lecturers encounter the independence of $(\overline{X}, \overline{Y})$ and S_p^2 when teaching the fact that the test statistic

$$\frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

is Student distributed with $m + n - 2$ degrees of freedom.

The problem of independence also comes across in the theory of linear regression.

Consider a model of linear regression with one controlled variable x and one response variable Y_x. For a given sequence of values $x_1, \ldots, x_n$ the variables Y_{x_i} will be denoted by Y_i. The following assumptions are made:

i) Y_x is $N(\alpha + \beta x, \sigma^2)$ distributed.

ii)   Given the values $x_1, \ldots, x_n$ for x, the variables $Y_1, \ldots, Y_n$ are independent.

Under these conditions the variables

$$\widehat{\beta} = \frac{\sum_i (x_i - \overline{x}) Y_i}{\sum_i (x_i - \overline{x})^2} \quad \text{and} \quad \widehat{\alpha} = \widehat{\beta} - \overline{x}\,\overline{Y}$$

are unbiased estimators for β and α. Writing $\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} x_i$, the so called "Sum of Squares of Errors" can be expressed as

$$SSE = \sum_i (Y_i - \widehat{Y}_i)^2.$$

Theorem C: *The variables* $(\widehat{\alpha}, \widehat{\beta})$ *and SSE are statistically independent.*

Proof: Observe that $\widehat{\alpha}, \widehat{\beta}$ and $Y_i - \widehat{Y}_i$ are linear statistics in $Y_1, \ldots, Y_n$. Apply the main theorem to verify that $(\widehat{\alpha}, \widehat{\beta})$ and $(Y_1 - \widehat{Y}_1, \ldots, Y_n - \widehat{Y}_n)$ are independent variables. Then conclude that $(\widehat{\alpha}, \widehat{\beta})$ and SSE are independent.

Remark: The independence of $(\widehat{\alpha}, \widehat{\beta})$ and SSE plays a role when teaching that the test statistic

$$\frac{(\widehat{\beta} - \beta)\sqrt{(n-2)\sum_i (x_i - \overline{x})^2}}{\sqrt{SSE}}$$

is Student distributed with $n - 2$ degrees of freedom.

References

[1] T.W. Anderson; An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, New York (1984).

[2] H. Bauer; Probability Theory. Walter de Gruyter, Berlin (1996).

[3] P.J. Bickel & K.A. Doksum; Mathematical Statistics: Basic Ideas and Selected Topics. Holden-Day, San Francisco (1977).

[4] W. Feller; An Introduction to Probability Theory and Its Applications. John Wiley & Sons, New York (1971).

[5] J.E. Freund; Mathematical Statistics. Prentice Hall, London 1992.

[6] G.R. Grimmett & D.R. Stirzaker; Probability and Random Processes. Oxford University Press, Oxford 1992.

[7] W.R. Pestman; Mathematical Statistics, an introduction. Walter de Gruyter, Berlin 1998.

[8] J.A. Rice; Mathematical Statistics and Data Analysis. Wadsworth & Brooks, Belmont, California 1988.

[9] J. Shuster; A Simple Method of Teaching the Independence of $\overline{X}$ and S^2. The American Statistician, vol. 27, no. 1, 1973.

[10] S.S. Wilks; Mathematical Statistics. John Wiley & Sons, New York (1962).

Wiebe R. Pestman
Department of Mathematics
University of Nijmegen
Toernooiveld
6525 ED Nijmegen
The Netherlands