WIKIPEDIA

Cochran's theorem

In statistics, **Cochran's theorem**, devised by William G. Cochran,[1] is a theorem used to justify results relating to the probability distributions of statistics that are used in the analysis of variance.[2]

Contents

Statement

Suppose $U_1, ..., U_N$ are i.i.d. standard normally distributed random variables, and there exist matrices $B^{(1)}, B^{(2)}, \ldots, B^{(k)}$, with $\sum_{i=1}^{k} B^{(i)} = I_N$. Further suppose that $r_1 + \cdots + r_k = N$, where r_i is the rank of $B^{(i)}$. If we write

$$Q_i = \sum_{j=1}^{N} \sum_{\ell=1}^{N} U_j B_{j,\ell}^{(i)} U_\ell$$

so that the Q_i are quadratic forms, then **Cochran's theorem** states that the Q_i are independent, and each Q_i has a chi-squared distribution with r_i degrees of freedom.[1]

Less formally, it is the number of linear combinations included in the sum of squares defining Q_i, provided that these linear combinations are linearly independent.

Proof

We first show that the matrices $B^{(i)}$ can be simultaneously diagonalized and that their non-zero eigenvalues are all equal to +1. We then use the vector basis that diagonalize them to simplify their characteristic function and show their independence and distribution.[3]

Each of the matrices $B^{(i)}$ has rank r_i and thus r_i non-zero eigenvalues. For each i, the sum $C^{(i)} \equiv \sum_{j \neq i} B^{(j)}$ has at most rank $\sum_{j \neq i} r_j = N - r_i$. Since $B^{(i)} + C^{(i)} = I_{N \times N}$, it follows that $C^{(i)}$ has exactly rank $N - r_i$.

Therefore $B^{(i)}$ and $C^{(i)}$ can be simultaneously diagonalized. This can be shown by first diagonalizing $B^{(i)}$. In this basis, it is of the form:

$$\begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & \cdots & & 0 \\ 0 & \lambda_2 & 0 & \cdots & \cdots & & 0 \\ 0 & 0 & \ddots & & & & \vdots \\ \vdots & \vdots & & \lambda_{r_i} & & & \\ \vdots & \vdots & & & 0 & & \\ 0 & \vdots & & & & \ddots & \\ 0 & 0 & \ldots & & & & 0 \end{bmatrix}.$$

Thus the lower $(N - r_i)$ rows are zero. Since $C^{(i)} = I - B^{(i)}$, it follows that these rows in $C^{(i)}$ in this basis contain a right block which is a $(N - r_i) \times (N - r_i)$ unit matrix, with zeros in the rest of these rows. But since $C^{(i)}$ has rank $N - r_i$, it must be zero elsewhere. Thus it is diagonal in this basis as well. It follows that all the non-zero eigenvalues of both $B^{(i)}$ and $C^{(i)}$ are +1. Moreover, the above analysis can be repeated in the diagonal basis for $C^{(1)} = B^{(2)} + \sum_{j>2} B^{(j)}$. In this basis $C^{(1)}$ is the identity of an $(N - r_1) \times (N - r_1)$ vector space, so it follows that both $B^{(2)}$ and $\sum_{j>2} B^{(j)}$ are simultaneously diagonalizable in this vector space (and hence also together $B^{(1)}$). By iteration it follows that all B-s are simultaneously diagonalizable.

Thus there exists an orthogonal matrix S such that for all i, $S^{\mathrm{T}} B^{(i)} S \equiv B^{(i)\prime}$ is diagonal, where any entry $B^{(i)\prime}_{x,y}$ is equal to 1 for $\sum_{j=1}^{i-1} r_j > x = y \leq \sum_{j=1}^{i} r_j$ and is equal to 0 for any other indices.

Let U_i' denote some specific linear combination of all U_i after transformation by S. Note that $\sum_{i=1}^{N} (U_i')^2 = \sum_{i=1}^{N} U_i^2$ due to the length preservation of the orthogonal matrix S.

The characteristic function of Q_i is:

$$\begin{aligned}
\varphi_i(t) &= (2\pi)^{-N/2} \int du_1 \int du_2 \cdots \int du_N e^{itQ_i} \cdot e^{-u_1^2/2} \cdot e^{-u_2^2/2} \cdots e^{-u_N^2/2} \\
&= (2\pi)^{-N/2} \left(\prod_{j=1}^{N} \int du_j \right) e^{itQ_i} \cdot e^{-\sum_{j=1}^{N} u_j^2/2} \\
&= (2\pi)^{-N/2} \left(\prod_{j=1}^{N} \int du_j' \right) e^{it \cdot \sum_{m=r_1+\cdots+r_{i-1}+1}^{r_1+\cdots+r_i} (u_m')^2} \cdot e^{-\sum_{j=1}^{N} {u_j'}^2/2} \\
&= (2\pi)^{-N/2} \left(\int e^{u^2(it-\frac{1}{2})} du \right)^{r_i} \left(\int e^{-\frac{u^2}{2}} du \right)^{N-r_i} \\
&= (1 - 2it)^{-r_i/2}
\end{aligned}$$

This is the Fourier transform of the chi-squared distribution with r_i degrees of freedom. Therefore this is the distribution of Q_i.

Moreover, the characteristic function of the joint distribution of all the Q_is is:

$$\begin{aligned}
\varphi(t_1, t_2, \ldots, t_k) &= (2\pi)^{-N/2} \left(\prod_{j=1}^{N} \int dU_j \right) e^{i \sum_{i=1}^{k} t_i \cdot Q_i} \cdot e^{-\sum_{j=1}^{N} U_j^2/2} \\
&= (2\pi)^{-N/2} \left(\prod_{j=1}^{N} \int dU_j' \right) e^{i \cdot \sum_{i=1}^{k} t_i \sum_{k=r_1+\cdots+r_{i-1}+1}^{r_1+\cdots+r_i} (U_k')^2} \cdot e^{-\sum_{j=1}^{N} U_j'^2/2} \\
&= (2\pi)^{-N/2} \prod_{i=1}^{k} \left(\int e^{u^2 (it_i - \frac{1}{2})} du \right)^{r_i} \\
&= \prod_{i=1}^{k} (1 - 2it_i)^{-r_i/2} = \prod_{i=1}^{k} \varphi_i(t_i)
\end{aligned}$$

From this it follows that all the Q_is are independent.

Examples

Sample mean and sample variance

If X_1, ..., X_n are independent normally distributed random variables with mean μ and standard deviation σ then

$$U_i = \frac{X_i - \mu}{\sigma}$$

is standard normal for each i. It is possible to write

$$\sum_{i=1}^{n} U_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{\sigma} \right)^2 + n \left(\frac{\overline{X} - \mu}{\sigma} \right)^2$$

(here $\overline{X}$ is the sample mean). To see this identity, multiply throughout by σ^2 and note that

$$\sum (X_i - \mu)^2 = \sum (X_i - \overline{X} + \overline{X} - \mu)^2$$

and expand to give

$$\sum (X_i - \mu)^2 = \sum (X_i - \overline{X})^2 + \sum (\overline{X} - \mu)^2 + 2 \sum (X_i - \overline{X})(\overline{X} - \mu).$$

The third term is zero because it is equal to a constant times

$$\sum (\overline{X} - X_i) = 0,$$

and the second term has just n identical terms added together. Thus

$$\sum (X_i - \mu)^2 = \sum (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2,$$

and hence

$$\sum \left(\frac{X_i-\mu}{\sigma}\right)^2 = \sum \left(\frac{X_i-\overline{X}}{\sigma}\right)^2 + n\left(\frac{\overline{X}-\mu}{\sigma}\right)^2 = Q_1+Q_2.$$

Now the rank of $B^{(2)}$ is just 1 (it is the square of just one linear combination of the standard normal variables). The rank of $B^{(1)}$ can be shown to be $n - 1$, and thus the conditions for Cochran's theorem are met.

Cochran's theorem then states that Q_1 and Q_2 are independent, with chi-squared distributions with $n - 1$ and 1 degree of freedom respectively. This shows that the sample mean and sample variance are independent. This can also be shown by Basu's theorem, and in fact this property *characterizes* the normal distribution – for no other distribution are the sample mean and sample variance independent.[4]

Distributions

The result for the distributions is written symbolically as

$$\sum \left(X_i-\overline{X}\right)^2 \sim \sigma^2 \chi^2_{n-1}.$$
$$n(\overline{X}-\mu)^2 \sim \sigma^2 \chi^2_1,$$

Both these random variables are proportional to the true but unknown variance σ^2. Thus their ratio does not depend on σ^2 and, because they are statistically independent. The distribution of their ratio is given by

$$\frac{n\left(\overline{X}-\mu\right)^2}{\frac{1}{n-1}\sum\left(X_i-\overline{X}\right)^2} \sim \frac{\chi^2_1}{\frac{1}{n-1}\chi^2_{n-1}} \sim F_{1,n-1}$$

where $F_{1,n-1}$ is the F-distribution with 1 and $n - 1$ degrees of freedom (see also Student's t-distribution). The final step here is effectively the definition of a random variable having the F-distribution.

Estimation of variance

To estimate the variance σ^2, one estimator that is sometimes used is the maximum likelihood estimator of the variance of a normal distribution

$$\widehat{\sigma}^2 = \frac{1}{n}\sum\left(X_i-\overline{X}\right)^2.$$

Cochran's theorem shows that

$$\frac{n\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

and the properties of the chi-squared distribution show that

$$E\left(\frac{n\widehat{\sigma}^2}{\sigma^2}\right) = E\left(\chi^2_{n-1}\right)$$
$$\frac{n}{\sigma^2}E\left(\widehat{\sigma}^2\right) = (n-1)$$
$$E\left(\widehat{\sigma}^2\right) = \frac{\sigma^2(n-1)}{n}$$

Alternative formulation

The following version is often seen when considering linear regression.[5] Suppose that $Y \sim N_n(0, \sigma^2 I_n)$ is a standard multivariate normal random vector (here I_n denotes the *n*-by-*n* identity matrix), and if $A_1, \ldots, A_k$ are all *n*-by-*n* symmetric matrices with $\sum_{i=1}^{k} A_i = I_n$. Then, on defining $r_i = \mathbf{Rank}(A_i)$, any one of the following conditions implies the other two:

- $\sum_{i=1}^{k} r_i = n,$
- $Y^T A_i Y \sim \sigma^2 \chi^2_{r_i}$ (thus the A_i are positive semidefinite)
- $Y^T A_i Y$ is independent of $Y^T A_j Y$ for $i \neq j.$

See also

- Cramér's theorem, on decomposing normal distribution
- Infinite divisibility (probability)

References

1. Cochran, W. G. (April 1934). "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance". *Mathematical Proceedings of the Cambridge Philosophical Society*. **30** (2): 178–191. doi:10.1017/S0305004100016595 (https://doi.org/10.1017%2FS0305004100016595).
2. Bapat, R. B. (2000). *Linear Algebra and Linear Models* (Second ed.). Springer. ISBN 978-0-387-98871-9.
3. Craig A. T. (1938) "On The Independence of Certain Estimates of Variances." *Annals of Mathematical Statistics*. 9, pp. 48–55
4. Geary, R.C. (1936). "The Distribution of "Student's" Ratio for Non-Normal Samples". *Supplement to the Journal of the Royal Statistical Society*. **3** (2): 178–184. doi:10.2307/2983669 (https://doi.org/10.2307%2F2983669). JFM 63.1090.03 (https://zbmath.org/?format=complete&q=an:63.1090.03). JSTOR 2983669 (https://www.jstor.org/stable/2983669).
5. "Cochran's Theorem (A quick tutorial)" (http://yangfeng.hosting.nyu.edu//slides/cochran's-theorem.pdf) (PDF).