

On Statistics Independent of a Sufficient Statistic

E. J. Williams

Institute of Statistics

Mimeo Series No. 176

February, 1957

ON STATISTICS INDEPENDENT OF A SUFFICIENT STATISTIC

by

E. J. Williams

North Carolina State College

1. Introduction. The purpose of this note is to prove a theorem on the independence of a sufficient statistic and other statistics derived from the same sample. A number of special cases of this theorem, all based on the normal or gamma (χ^2) distributions, have been published from time to time; but even these results do not seem to be well known. We confine ourselves to distributions depending on only one unknown parameter, although certain generalizations to more than one parameter can readily be made.

The theorem has already been proved by Basu (1955), but using more elaborate mathematical techniques than will be found here.

As we remark below, we have not found any applications of the theorem to distributions other than the normal and gamma distributions, but it is conceivable that such applications may arise in investigations of sampling distributions.

2. The Main Result. For a sample drawn from a population of values of x with distribution depending on a parameter θ , let z be a statistic sufficient for θ , and g any statistic whose distribution is independent of θ . Then g and z are independently distributed.

Before proving this result, we note that the converse is easily established. For if g and z are independent, the conditional distribution of g for fixed z is the same as the unconditional distribution; but the conditional distribution of any statistic, for fixed z , is independent of θ , so the distribution of g is independent of θ .

The main result is proved as follows. Since a sufficient statistic for θ

exists, the distribution function of an individual value must be of the form (Koopman 1936; Pitman 1936)

$$e^{A(\theta)B(x) + C(\theta) + D(x)}$$

where $A(\theta)$ and $C(\theta)$ are functions only of the parameter, and $B(x)$ and $D(x)$ are functions only of the sample values.

Then the joint density of a random sample x_1, x_2, \dots, x_n is

$$e^{A(\theta) \sum_i B(x_i) + n C(\theta) + \sum_i D(x_i)}$$

The sufficient statistic is any function of $\sum_i B(x_i)$; without loss of generality we shall take

$$z = \sum_i B(x_i),$$

and for convenience write

$$\sum_i D(x_i) = X.$$

We shall also transform the parameter by denoting the coefficient of z by θ ; it is not necessary to the proof that this transformation be possible, but it simplifies the discussion. The joint density is now

$$e^{\theta z + n C(\theta) + X}$$

The joint characteristic function of z and g is

$$\begin{aligned} \phi(t_1, t_2) &= E(e^{i(t_1 z + t_2 g)}) \\ &= \int \dots \int e^{it_1 z + it_2 g + \theta z + nC(\theta) + X} dx_1 \dots dx_n \end{aligned}$$

The characteristic function of z is

$$\begin{aligned}\phi_1(t_1) &= \int \dots \int e^{it_1 z + \theta z + nC(\theta) + X} dx_1 \dots dx_n \\ &= e^{-n[C(\theta+it_1) - C(\theta)]}\end{aligned}$$

This last result is well known (Fisher 1934).

Now the characteristic function of g is

$$\phi_2(t_2) = \int \dots \int e^{it_2 g + \theta z + nC(\theta) + X} dx_1 \dots dx_n.$$

Since g is distributed independently of θ , ϕ_2 is independent of θ , and so is unaltered if θ is replaced by $\theta+it_1$.

$$\begin{aligned}\phi_2(t_2) &= \int \dots \int e^{it_2 g + it_1 z + \theta z + nC(\theta+it_1) + X} dx_1 \dots dx_n \\ &= e^{n[C(\theta+it_1) - C(\theta)]} \phi(t_1, t_2) \\ &= \phi(t_1, t_2) / \phi_1(t_1) .\end{aligned}$$

Hence

$$\phi(t_1, t_2) = \phi_1(t_1) \phi_2(t_2) ,$$

so that g and z are independently distributed.

3. Applications. This general result has been repeatedly applied to the normal and gamma distributions.

(i) If θ is the mean of a normal population, the sample mean \bar{x} is a sufficient statistic for θ . Since θ is a location parameter (i.e., the density of x may be written in the form $f(x-\theta)$), any statistic independent of location is independent of θ and hence of \bar{x} . This result was apparently first published by Daly (1946), though it was known earlier.

Thus the sample variance s^2 , the mean square successive difference δ^2 , the range w , and other statistics used to estimate the variance of the distribution, are all distributed independently of \bar{x} (see Geary (1936), Lukacs (1942), Geisser (1956)). Other more complicated examples are easily constructed.

(ii) If θ^{-1} is the variance of a normal population with zero mean, the sample sum of squares S is a sufficient statistic for θ . The density of S for a sample of n is

$$\frac{1}{\Gamma(\frac{n}{2})} \left(\frac{\theta}{2}\right)^{\frac{n}{2}} e^{-\frac{\theta S}{2}} S^{\frac{n}{2}-1} dS ;$$

$\theta S/2$ is said to have a $\Gamma(\frac{n}{2})$ distribution, or θS to be distributed as χ^2 with n degrees of freedom.

Since θ^{-1} is a scale parameter for the normal population, any statistic independent of scale (i.e., homogeneous of degree zero in the x) is independent of θ and hence of S .

Again, if samples of size n_1, n_2, \dots, n_m are drawn from the population, yielding sums of squares S_1, S_2, \dots, S_m , the sufficient statistic for θ is

$$S = S_1 + S_2 + \dots + S_m$$

with degrees of freedom

$$n = n_1 + n_2 + \dots + n_m .$$

Any function of the S_i homogeneous of degree zero is independent of S . This result for gamma distributions was first given by Pitman (1937).

Various special cases of this result have been either established or implied; the following statistics have been shown to be distributed independently of the sample variance:

sample skewness, kurtosis and other measures of departure from normality (Fisher 1930);

ratio of mean square successive difference to the variance (von Neumann 1941);

serial correlation coefficients (Koopmans 1942).

4. Limitations. It will be noted that, in the above applications, the parameters are either location or scale parameters. In such cases it is easy to derive statistics whose distributions are independent of the parameter. However we know of no other cases in which a statistic distributed independently of a parameter may be derived.

Owen (1948) has investigated this question thoroughly in connection with a method of estimation proposed by Fisher (1936). However, in all the examples he gives of statistics distributed independently of population parameters, the parameter may be expressed as parameters of location or scale. For instance, the example he gives on page 9 (with notation modified slightly) is the joint normal density

$$\frac{1}{2\pi} e^{-\frac{1}{2}[(x-\theta)^2 + (y-\phi)^2]}$$

where the statistics θ , ϕ are connected by the relationship

$$(\theta-a)^2 + (\phi-b)^2 = r^2$$

a , b and r being known constants. He shows that

$$S^2 = (x-a)^2 + (y-b)^2$$

has a distribution independent of the unknown parameter. Now if we write

$$\begin{aligned} \theta &= a+r \cos \psi & , & & \phi &= b+r \sin \psi \\ x &= a+S \cos u & , & & y &= b+S \sin u & , \end{aligned}$$

the density becomes

$$\frac{1}{2\pi e} e^{-\frac{1}{2}(S^2 - 2rS \cos(u-\psi) + r^2)} S$$

This representation shows that ψ , which is equivalent to the unknown parameter, is a "location parameter" for the angle u .

If we are prepared to assume, then, that only for very exceptional distributions, if any, will parameters other than location and scale parameters have statistics independent of them, we see that the theorem has very limited application indeed. In fact, it is limited to distributions whose location and scale parameters admit sufficient statistics.

Now it is readily established that, of distributions whose end-points are independent of θ (and hence are $\pm \infty$), only the normal distribution admits a statistic sufficient for location; for the log density

$$L = \theta z + nC(\theta) + x$$

must also be of the form

$$\sum f(x_i - \theta).$$

Comparison of the two forms shows that f must be a quadratic function, and that $\sum x$ is the sufficient statistic; these properties characterize the normal distribution. Of distributions whose end-points depend on θ , only the exponential, with density

$$\begin{array}{ll} \lambda e^{-\lambda(x-\theta)} & x \geq \theta \\ 0 & x < \theta \end{array} \quad (\lambda > 0)$$

admits a sufficient statistic for location; the statistic is the smallest member of the sample (Pitman 1937).

In order that a distribution admit a sufficient statistic for the scale

parameter, which we denote by θ^{-1} , we require the equivalence of

$$L = \theta z + nC(\theta) + X$$

and

$$\sum f(\theta x_i) + n \log \theta$$

Now $f(\theta x)$ may contain, apart from a constant, only

- (i) a linear term, summing to θz ,
- and (ii) a logarithmic term, breaking up into a term in θ and a term in x .

Hence the most general form is

$$f(\theta x) = -\theta x + \left(\frac{k}{2} - 1\right)(\log \theta + \log x) + \text{constant}.$$

This characterizes the gamma distribution, or χ^2 distribution with k degrees of freedom.

Thus the theorem is remarkable in that, although it appears to be a generalization of known results, its scope of application is limited and it does not effectively add to what is already known. It is possible, however, that further investigations of sampling distributions will reveal wider applications of the theorem.

5. Connection with estimation theory. Fisher (1934, 1936) has shown that there are two situations in which a parameter may be estimated exhaustively by means of a single statistic. One case is that in which the statistic is sufficient, so that the dependence of the likelihood of the original sample on the parameter is equivalent to the dependence of the statistic on the parameter.

The other case is that in which, although no sufficient statistic exists, it is possible to determine, for a sample of n , a set of $n-1$ functionally independent statistics whose joint distribution is independent of the parameter. This set of statistics defines the configuration of the sample. The set of all

possible samples with the same configuration has but one degree of freedom; and if, as is often the case, it is appropriate to confine attention to the restricted set with the same configuration as the observed sample, we then have a one-to-one correspondence between the values of the parameter and the values of any statistic chosen to estimate it. The distribution of this statistic, conditional on the sample configuration, provides, in conjunction with the configuration, a means of making exhaustive estimates. Fisher (1934) has shown the application of this method to the estimation of location and scale parameters, and Pitman (1938, 1939) has elaborated these ideas and obtained a number of general results.

As Welch (1939) has pointed out, use of the conditional distribution may lead on the average to a wider confidence range for the parameter, and hence less efficient estimation, than would the unconditional distribution of some statistic. Nevertheless, in many cases information will be needed on the configuration of the sample as well as the estimates which it yields, and in such cases the conditional distribution will be appropriate.

From the discussion in this section it will be clear that the theorem given in this paper makes no direct contribution to the theory of estimation. For if a distribution admits a sufficient statistic z , the properties of other statistics (independence of z , etc.) are not relevant for estimation purposes. Furthermore, aspects of the sample other than the estimation of θ would be examined by means of the distribution of the sample conditional on z . On the other hand, if the parameter is one of location or scale, corresponding to which statistics independent of the parameter can readily be found, a sufficient statistic exists, and the theorem is applicable, only in the well known examples of normal, exponential and gamma distributions.

Any value the theorem may have would be in establishing the independence of two statistics under certain conditions. This result is frequently helpful in simplifying the study of distributions.

REFERENCES

- Basu, D. (1955). On statistics independent of a complete sufficient statistic. Sankhya 15, 377-380.
- Daly, J.F. (1946). On the use of the sample range in an analogue of Student's t-test. Ann. Math. Stat. 17, 71-74.
- Fisher, R.A. (1930). The moments of the distribution for normal samples of measures of departure from normality. Proc. Roy. Soc. A 130, 16-28.
- Fisher, R.A. (1934). Two new properties of mathematical likelihood. Proc. Roy. Soc. A 144, 285-307.
- Fisher, R.A. (1936). Uncertain inference. Proc. Amer. Acad. Arts and Sciences 71, 245-258.
- Geary, R.C. (1936). Distribution of "Students'" ratio for non-normal samples. J. Roy. Stat. Soc. Supp. 3, 178-184.
- Geisser, S. (1956). A note on the normal distribution. Ann. Math. Stat. 27, 858-859.
- Koopman, B.O. (1936). On distributions admitting a sufficient statistic. Trans. Amer. Math. Soc. 39, 399-409.
- Koopmans, T. (1942). Serial correlation and quadratic forms in normal variables. Ann. Math. Stat. 13, 14-33
- Lukacs, E. (1942). A characterization of the normal distribution. Ann. Math. Stat. 13, 91-93.
- Owen, A.R.G. (1948). Ancillary statistics and fiducial distributions. Sankhya 9,1,
- Pitman, E.J.G. (1936). Sufficient statistics and intrinsic accuracy. Proc. Camb. Phil. Soc. 32, 567-579.
- Pitman, E.J.G. (1937). The "closest" estimates of statistical parameters. Proc. Camb. Phil. Soc. 33, 212-222.
- Pitman, E.J.G. (1938). The estimation of the location and scale parameters of a continuous population of any given form. Biometrika 30, 391-421.
- Pitman, E.J.G. (1939). Tests of hypotheses concerning location and scale parameters Biometrika 31, 200-215.
- von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. Ann. Math. Stat. 12, 367-395.
- Welch, B.L. (1939). On confidence limits and sufficiency, with particular reference to parameters of location. Ann. Math. Stat. 10, 58-69.