A Useful Property of Some Symmetric Statistics

TIMOTHY C. BROWN and GEOFFREY K. EAGLESON*

The distributions of some statistics arising from the comparison of two populations share a common structure. Consequences of this structure are derived and their usefulness indicated. The examples of statistics to which the theory applies include various correlation coefficients.

KEY WORDS: Independence; Pairwise independence; Correlations.

1. INTRODUCTION

In problems comparing many samples, one frequently makes pairwise comparisons that are, of necessity, interdependent. In Section 2, a restriction on the distribution of the pairwise comparisons is defined (and called Property $*$). This Property is restrictive but carries with it a number of interesting and useful consequences, enabling one to approximate to the distribution of the maximum of the pairwise comparisons, among other things.

Surprisingly, many common statistics have Property $*$, rank and product moment correlation coefficients being the most frequently used examples (see Sec. 3). Furthermore, some of the important consequences of $*$ do not seem to have been explicitly noted, even for correlation coefficients.

Since Property $*$ involves the ideas of independence and conditional distributions, we hope this article will be useful to teachers of both probability and statistics. We also give a number of examples of triples of random variables that are pairwise independent but not totally independent. Although many other examples of this phenomenon exist (e.g., Bernstein (quoted by Tucker 1962, p. 20); Geisser and Mantel 1962; Pitman and Williams 1967; Feller 1968, p. 220; Wong 1972; Chung 1974, Section 3.3, Ex. 2; Bühler and Miescke 1981), the examples of Section 3 are very simple, not contrived, and have useful applications.

2. GENERALITIES

In statistics, we often have several random samples represented by random vectors $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}, \ldots$. Let us assume that under a null hypothesis these vectors are independent and identically distributed. Furthermore, let us suppose that we have statistics $g(\mathbf{X}, \mathbf{Y})$ to assess the null hypothesis for two samples $\mathbf{X}$ and $\mathbf{Y}$. We shall assume that the comparison g is symmetric.

In a surprisingly large number of cases (see Sec. 3), the conditional distribution of $g(\mathbf{X},\mathbf{Y})$, given $\mathbf{X} = \mathbf{x}$, does not depend on the value $\mathbf{x}$. Let us call such a condition Property $*$.

Property $*$ has many useful consequences, including the following:

1. The random variable $g(\mathbf{X}, \mathbf{Y})$ is independent of $\mathbf{X}$ and, by symmetry, $g(\mathbf{X}, \mathbf{Y})$ is also independent of $\mathbf{Y}$. Thus $\{\mathbf{X}, \mathbf{Y}, g(\mathbf{X}, \mathbf{Y})\}$ is a set of three random vectors that are independent in pairs, but not as a triple.
2. The random variables $\{g(\mathbf{X}, \mathbf{Y}), g(\mathbf{X}, \mathbf{Z}), g(\mathbf{Y}, \mathbf{Z})\}$ are also independent in pairs but will not usually be independent as a triple. In fact, conditional on $\mathbf{X} = \mathbf{x}$, $g(\mathbf{X}, \mathbf{Y})$ and $g(\mathbf{X}, \mathbf{Z})$ are independent, and their identical conditional distributions are, by supposition, the same as their unconditional distributions.
3. One often wishes to combine the g statistics from, say, samples $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ by using the maximum, M, of $g(\mathbf{X}, \mathbf{Y})$, $g(\mathbf{X}, \mathbf{Z})$, and $g(\mathbf{Y}, \mathbf{Z})$. The pairwise independence of Consequence 2 makes it easy to compute Bonferroni bounds (Feller 1968, p.110) for the tail of the distribution of M. Specifically,
$$3\alpha(1 - \alpha) \leq P(M \geq g_0) \leq 3\alpha,$$
where $\alpha = P(g(\mathbf{X}, \mathbf{Y}) \geq g_0)$ is the individual significance level.
4. There are, of course, other ways to combine the information contained in the g's from several samples. Some of these would involve linear combinations of functions of the g's. The pairwise independence makes it simple to compute the variance of such combinations. Specifically, for any real-valued functions a, b, and c,
$$\operatorname{var}[a(g(\mathbf{X}, \mathbf{Y})) + b(g(\mathbf{Y}, \mathbf{Z})) + c(g(\mathbf{X}, \mathbf{Z}))]$$
$$= \operatorname{var}[a(g(\mathbf{X}, \mathbf{Y}))] + \operatorname{var}[b(g(\mathbf{Y}, \mathbf{Z}))] + \operatorname{var}[c(g(\mathbf{X}, \mathbf{Z}))].$$
5. If the distribution of $\mathbf{X}$ is changed but that of $\mathbf{Y}$ is not, then the distribution of $g(\mathbf{X}, \mathbf{Y})$ remains the same.

3. EXAMPLES

3.1 Rank Correlation Coefficients

The null hypothesis is that k sets of ranks $\mathbf{R}_1$, $\mathbf{R}_2, \ldots, \mathbf{R}_k$ for n objects, labeled $1, 2, \ldots, n$, are independent and that each is uniformly distributed on the reorderings of $\mathbf{i} = (1, 2, \ldots, n)$. Two popular statistics for detecting departure from the independence of $\mathbf{R}_1$ and $\mathbf{R}_2$ are Spearman's rho, $\rho(\mathbf{R}_1, \mathbf{R}_2)$ (Lehmann 1975, p.300) and Kendall's tau, $\tau(\mathbf{R}_1, \mathbf{R}_2)$ (Lehmann 1975, p.316). In both cases, Property $*$ holds. To see this for

*Timothy C. Brown is a Lecturer with the Department of Mathematics, Monash University, Clayton, Victoria, 3168, Australia. Geoffrey K. Eagleson is Principal Research Scientist, CSIRO, Division of Mathematics and Statistics, Lindfield, New South Wales, 2070, Australia. The authors are grateful to John Best and Murray Cameron for many useful discussions, to Evan Williams for some references, and to the referees for some helpful comments.

ρ, let $\mathbf{r}$ be an arbitrary reordering of $\mathbf{i}$. Suppose that the objects are now relabeled so that the ranking $\mathbf{i}$ is changed to $\mathbf{r}$. Let $\mathbf{S}_1, \mathbf{S}_2$ be the rankings obtained from $\mathbf{R}_1, \mathbf{R}_2$ after the relabeling. Clearly $(\mathbf{S}_1, \mathbf{S}_2)$ has the same distribution as $(\mathbf{R}_1, \mathbf{R}_2)$, and so

$$\begin{aligned} P(\rho(\mathbf{R}_1, \mathbf{R}_2) = \rho_0 \mid \mathbf{R}_1 = \mathbf{r}) &= P(\rho(\mathbf{S}_1, \mathbf{S}_2) = \rho_0 \mid \mathbf{S}_1 = \mathbf{r}) \\ &= P(\rho(\mathbf{R}_1, \mathbf{R}_2) = \rho_0 \mid \mathbf{S}_1 = \mathbf{r}) \\ &= P(\rho(\mathbf{R}_1, \mathbf{R}_2) = \rho_0 \mid \mathbf{R}_1 = \mathbf{i}) \end{aligned} \tag{3.1}$$

by the definition of $\mathbf{S}_1$ and the fact that relabeling the objects does not change the value of ρ. An identical argument applies to τ.

The Consequences 1–5 of Property $*$ do not seem to have been explicitly noted for τ and ρ. As an illustration of Consequence 3, suppose that the $\binom{k}{2}$ different Spearman's rhos are calculated from k sets of ranks $\mathbf{R}_1, \ldots, \mathbf{R}_k$. The maximum of these coefficients is a useful statistic for detecting *any* dependence among the rankings. The Bonferroni bounds can be calculated for the tail of the distribution of the maximum. The probability that the maximum of the ρ's is larger than ρ_0 is bounded between

$$\binom{k}{2} P(\rho(\mathbf{R}_1, \mathbf{R}_2) \geq \rho_0) - \tfrac{1}{2}\binom{k}{2}\left(\binom{k}{2} - 1\right) P^2(\rho(\mathbf{R}_1, \mathbf{R}_2) \geq \rho_0)$$

and

$$\binom{k}{2} P(\rho(\mathbf{R}_1, \mathbf{R}_2) \geq \rho_0).$$

For large ρ_0 and moderate k, the Bonferroni bounds are quite close: if ρ_0 is the $100(1-\alpha)$ percentile of ρ and $k = 5$, then the upper bound is 10α, and the difference between the upper and lower bounds is $45\alpha^2$. For example, for $n = 7$ and $\alpha = .0062$, we have $\rho_0 = 25/28$, an upper bound of .062, and a difference of .0017. (Percentiles may be found, e.g., from Table N of Lehmann 1975.)

3.2 Product Moment Correlation Coefficients

If a vector of observations $\mathbf{X}_i = (X_{i1}, \ldots, X_{in})$ $(i = 1, \ldots, k)$ is standardized by setting

$$Y_{ij} = (X_{ij} - \overline{X}_i) \Big/ \left[\sum_{\alpha=1}^{n} (X_{i\alpha} - \overline{X}_i)^2\right]^{1/2},$$

where $\overline{X}_i$ is the arithmetic mean for $\mathbf{X}_i$, then $\mathbf{Y}_i$ represents a point on the intersection of the unit sphere in $\mathbf{R}^n$, and the hyperplane $\sum y_\alpha = 0$; that is, it lies on the unit sphere S_{n-2} in $\mathbf{R}^{n-1}$. The product moment correlation coefficient, $r(\mathbf{X}_1, \mathbf{X}_2)$, is then simply the cosine of the angle between $\mathbf{Y}_1$ and $\mathbf{Y}_2$, so that $r(\mathbf{X}_1, \mathbf{X}_2) = r(\mathbf{Y}_1, \mathbf{Y}_2) = \mathbf{Y}_1^T\mathbf{Y}_2$. In this case, Property $*$ holds if the X's are all independent and normally distributed. To see this, fix a direction $\mathbf{i}$ in S_{n-2} and let $\mathbf{y}$ be an arbitrary direction. Let $\mathbf{Z}_1, \mathbf{Z}_2$ be the result on $\mathbf{Y}_1, \mathbf{Y}_2$ of a rotation of S_{n-2} in such a way that $\mathbf{i}$ becomes $\mathbf{y}$. Since the $\mathbf{Y}$'s are uniform on S_{n-2} (Fisher 1915), the argument of (3.1) for ρ goes through with $\mathbf{Y}$'s replacing $\mathbf{R}$'s and $\mathbf{Z}$'s replacing $\mathbf{S}$'s (of course we must change some equalities to inequalities since the distribution of r is continuous).

The pairwise independence (Consequence 2) of the product moment correlation coefficients was shown by Geisser and Mantel (1962), in which an argument based on the joint distribution of the coefficients was outlined. Consequence 3 was also noted by Geisser and Mantel and applied to give the variance of the number of correlation coefficients among $r(\mathbf{X}_i, \mathbf{X}_j)$, $1 \leq i < j \leq k$, which are significant at, say, the 5% level. Moran (1980) notes Consequence 4 for the maximum of the $r(\mathbf{X}_i, \mathbf{X}_j)$, $1 \leq i < j \leq k$. We were unable to find Consequence 5 explicitly stated, but it seems to be of particular practical importance.

3.3 Distances on Circles, Toruses, and Spheres

The argument of Section 3.2 works equally well if we consider $d(\mathbf{Y}_1, \mathbf{Y}_2)$, the distance between $\mathbf{Y}_1$ and $\mathbf{Y}_2$, or if the $\mathbf{Y}$'s are on a torus or a circle. The last case is particularly elementary and here $d(\mathbf{Y}_1, \mathbf{Y}_2)$, $d(\mathbf{Y}_1, \mathbf{Y}_3)$, $d(\mathbf{Y}_2, \mathbf{Y}_3)$ provide a simple but eminently practical example of random variables independent in pairs but not as a triple. The title of this section is the title of a paper by Silverman (1978), who notes Consequence 2 and proves some asymptotic results. The last section of Silverman (1978) provides a general framework, in terms of a group of transformations on an abstract space, for all of the examples in this article. We note that the argument for Property $*$ in 3.1 and 3.2 easily generalizes to this framework and thus provides an alternative to Silverman's derivation of Consequence 2.

3.4 Interaction in Spatial Patterns

Consider k binary classifications of n objects labeled $1, \ldots, n$. Record the results as n vectors $\mathbf{P}_1, \ldots, \mathbf{P}_k$ of 0's and 1's. Under the null hypothesis, we suppose that the vectors are independent and that, if m is the number of 1's in each vector, $\mathbf{P}_i$ is uniform on $\{\mathbf{i} : i_j = 0 \text{ or } 1, \sum i_j = m\}$. Then $\mathbf{P}_1^T\mathbf{P}_2$ measures the extent to which the 1 values of $\mathbf{P}_1$ are attributed to the same objects as in $\mathbf{P}_2$. This statistic is invariant under relabelings of the objects and so is the distribution of each $\mathbf{P}_i$. Thus, analogous arguments to those of Section 3.1 show that Property $*$ holds. Consequences 2 and 4 in this example were noted by Geisser and Mantel (1962). Moreover, the statistic could be useful in testing whether patterns of points of different types interact. One could place a grid of n cells on the space supporting the pattern. Then $\mathbf{P}_{il}$ would record the presence or absence of points of type i in cell l. If two patterns attract each other $\mathbf{P}_1^T\mathbf{P}_2$ should be large, while if they repel each other $\mathbf{P}_1^T\mathbf{P}_2$ should be small.

We note that with slight modifications, Consequences 1–5 also hold in the more practical case in which the numbers of points of different types vary.

[*Received January 1983. Revised August 1983.*]

REFERENCES

BÜHLER, W.J., and MIESCKE, K.J. (1981), "On $(n - 1)$-wise and Joint Independence and Normality of n Random Variables: An Example," *Communications in Statistics—Theory and Methods,* 10, 927–930.

CHUNG, K.L. (1974), *A Course in Probability Theory,* 2nd ed., New York: Academic Press.

FELLER, W. (1957), *An Introduction to Probability Theory and Its Applications,* 2nd ed., New York: John Wiley.

FISHER, R. (1915), "Frequency Distribution of the Values of the Correlation Coefficient in a Sample of Values From an Indefinitely Large Population," *Biometrika,* 10, 507–521.

GEISSER, S., and MANTEL, N. (1962), "Pairwise Independence of Jointly Dependent Variables," *Annals of Mathematical Statistics,* 33, 290–291.

LEHMANN, E.L. (1975), *Nonparametrics, Statistical Methods Based on Ranks,* San Francisco: Holden-Day.

MORAN, P.A.P. (1980), "Testing the Largest of a Set of Correlation Coefficients," *Australian Journal of Statistics,* 22, 289–297.

PITMAN, E.J.G., and WILLIAMS, E.G. (1967), "Cauchy-Distributed Functions of Cauchy Variates," *Annals of Mathematical Statistics,* 38, 916–918.

SILVERMAN, B. (1978), "Distance on Circles, Toruses, and Spheres," *Journal of Applied Probability,* 15, 136–143.

TUCKER, H.E. (1962), *An Introduction to Probability and Mathematical Statistics,* New York: Academic Press.

WONG, C.K. (1972), "Note on Mutually Independent Events," *The American Statistician,* 26, 27–28.