Math. Proc. Camb. Phil. Soc. (1975), **77**, 185
With 1 *text-figure*
MPCPS 77–15
Printed in Great Britain

Dissociated random variables

BY W. G. McGINLEY AND R. SIBSON
University of Cambridge

(*Received* 27 *June* 1974)

We introduce the idea of dissociatedness for multiply-subscripted random variables: this is a weaker property than independence in general, coinciding with it in the singly-subscripted case; but it is still strong enough to allow a suitable version of the Strong Law of Large Numbers to hold. This result can be applied to prove a weak convergence result of interest in data analysis.

1. *Colouring mosaics.* We define a *mosaic* to be an $l \times j$ block of cells. A *colouring* of a mosaic with n colours is an assignment of a colour to each cell; it is not necessary for all the colours to be used, and they may be repeated, so the total number of colourings of an $l \times j$ mosaic with n colours is n^{lj}. A colouring is *endochrome* if there is some row of the mosaic such that the set of colours used for that row is disjoint from the set of colours used elsewhere in the mosaic; a colouring which is not endochrome is *exochrome*. If n is large, clearly most of the colourings will be endochrome: we calculate the leading power of n in the expression for the number of exochrome colourings. The cases l even and l odd exhibit slightly different behaviour.

Counting Lemma. The number of exochrome colourings of an $l \times j$ mosaic with n colours is:

(i) $$\frac{j^{2m}.(2m)!}{2^m.m!} n^{2mj-m} + O(n^{2mj-m-1}) \quad \text{if} \quad l = 2m;$$

(ii) $$\left\{\frac{j^{2m+1}.(j-1).(2m+1)!}{2^m.(m-1)!} + \frac{j^{2m+1}.(2m+1)!}{2^m.3.(m-1)!}\right\} \times n^{(2m+1)j-(m+1)} + O(n^{(2m+1)j-(m+1)-1}) \quad \text{if} \quad l = 2m+1.$$

Proof. (i) $l = 2m$. The most polychrome colourings which are exochrome are of the form

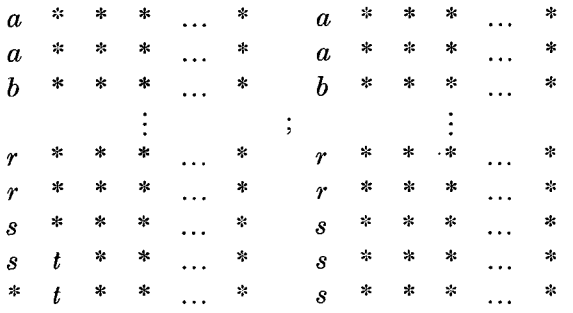
$$\begin{array}{cccccc} a & * & * & * & \dots & * \\ a & * & * & * & \dots & * \\ b & * & * & * & \dots & * \\ b & * & * & * & \dots & * \\ & & & \vdots & & \\ r & * & * & * & \dots & * \\ r & * & * & * & \dots & * \\ s & * & * & * & \dots & * \\ s & * & * & * & \dots & * \end{array}$$

and other similar patterns obtained by changing the pairing of the rows and the positions of the paired cells within each row of the mosaic. Cells which may be coloured

https://doi.org/10.1017/S0305004100049513 Published online by Cambridge University Press

arbitrarily are marked *, and paired cells which must be the same colour as one another are lettered; cells which are not explicitly paired with one another are all differently coloured. The leading term is generated by such colourings, and the usual combinatorial counting arguments give the coefficient stated.

(ii) $l = 2m+1$. The argument is similar, but two terms must be considered, arising from colourings of the following forms:



Corollary. The number of exochrome colourings of a $4 \times j$ *mosaic is* $O(n^{4j-2})$.

The corollary is the version of the counting lemma which we shall actually use. Other versions are sometimes useful in probability theory, particularly the case $l = 2m, j = 1$.

2. *A strong law for dissociated random variables.* If x and y are j-vectors, we write $x \leqslant y$ to mean $x_i \leqslant y_i$ for $i = 1, \ldots, j$. (ξ) will denote the j-vector all of whose entries are equal to ξ.

Let $\{X_r\}$ be a family of random variables subscripted by j-vectors of integers. We call such a family *dissociated* if, loosely speaking, members of the family can only be dependent by having a common subscript. A precise definition is that $\{X_r\}$ is dissociated if for every two subfamilies A and B of $\{X_r\}$ such that the set of integers which occur in subscripts of members of A is disjoint from the corresponding set for B, the joint distributions of the subfamilies A and B are independent of one another. A simple example is as follows. Let $Y_1, \ldots, Y_n$ be independent random variables, and define X_r by

$$X_r = \max\{Y_{r_1}, Y_{r_2}, \ldots, Y_{r_j}\}.$$

The family $\{X_r\}$ is dissociated. An independent family is dissociated, but, as the example shows, the converse is false in general unless $j = 1$.

It is natural to attempt to prove a strong law of large numbers for a dissociated family, but some care is needed in framing this: whereas it may be reasonable to assume that $E(X_r)$ is constant for all r such that $r_1, \ldots, r_j$ are distinct, there is no reason why the same should be true when r contains duplicated subscript entries; reference to the example makes the point clear. Thus we shall avoid this case – which in applications it is easy to do – and work with an *off-diagonal* dissociated family subscripted by integer vectors r with $r_1, \ldots, r_j$ all distinct.

Theorem [*Off-diagonal SLLN*]. *Let* $\{X_r\}$ *be an off-diagonal dissociated family such that* $E(X_r) = 0$ *for all* r *and there exists* $K < \infty$ *such that* $E(X_r^4) \leqslant K$ *for all* r. *Define* $S_n = \Sigma\{X_r : (1) \leqslant r \leqslant (n)\}$. *Then* $S_n/N \to 0$ *almost surely as* $n \to \infty$, *where* $N = n(n-1)\ldots(n-j+1)$.

https://doi.org/10.1017/S0305004100049513 Published online by Cambridge University Press

Proof. $E(S_n^4)$ consists of N^4 terms, and by dissociatedness all these terms are zero except for those where the integers $1, \ldots, n$ form an exochrome colouring of the $4 \times j$ mosaic of subscripts in the term; the non-vanishing terms are all bounded by K. Thus $E(S_n^4) = O(n^{4j-2})$, and the proof can then be completed by using the usual arguments based on Chebyshev's inequality and the first Borel–Cantelli lemma. It is, of course, purely for aesthetic reasons that we have divided by the number N of terms being summed; n^j would do just as well.

The application which motivated us to investigate the above result in fact involves only uniformly bounded random variables, so trivially we have $E(X_r^4) \leqslant K$. We have not investigated how far, if at all, this assumption is unnecessary, but it is to be expected that some price must be paid for weakening independence to dissociatedness.

3. *Almost sure convergence in distribution.* If $\{x_1, \ldots, x_N\}$ is a set of real numbers, we define the *empirical distribution function* (*EDF*) of $x_1, \ldots, x_N$ to be

$$F(x) = N^{-1} \sum_1^N \Delta(x - x_i),$$

where Δ is the step-function

$$\begin{aligned} \Delta(\xi) &= 0 \quad \text{if} \quad \xi < 0, \\ &= 1 \quad \text{if} \quad \xi \geqslant 0. \end{aligned}$$

Let $(\Omega, \mathscr{F}, P)$ be a probability space, and let f: $(\Omega, \mathscr{F})^j \to (R, \mathscr{B})$ be measurable; that is, f is a real statistic of j observations. Let F be the DF of f when $(\Omega, \mathscr{F})^j$ has the probability measure P^j. If we take a large number n of independent observations from $(\Omega, \mathscr{F}, P)$, divide these into disjoint subsets of size j, and calculate the value of f on these subsets, then of course as $n \to \infty$, the EDF of the resultant values almost surely converges in distribution to F. We show that this is still true if the values of f are calculated on all subsets, not just non-overlapping ones.

Theorem [*Convergence of Sample Distributions*]. *Let* $(\Omega, \mathscr{F}, P), f, F$ *be as above. Let* $\omega_1, \ldots, \omega_n$ *be independent observations from* $(\Omega, \mathscr{F}, P)$, *and let* F_N *be the EDF of the* $N = n(n-1)\ldots(n-j+1)$ *values of* f *calculable using exactly* j *of* $\omega_1, \ldots, \omega_n$ *at a time in all possible orders. Then* $F_N \to F$ *in distribution, almost surely as* $n \to \infty$.

Proof. Define X_r by

$$\begin{aligned} X_r(\omega) &= 0 \quad \text{if} \quad f(\omega_{r_1}, \ldots, \omega_{r_j}) < \alpha, \\ &= 1 \quad \text{otherwise.} \end{aligned}$$

Then $E(X_r) = F(\alpha)$, and $\{X_r - F(\alpha)\}$ is a dissociated off-diagonal family of bounded random variables with zero expectation. It follows from the off-diagonal SLLN that $F_N(\alpha) \to F(\alpha)$ almost surely as $n \to \infty$. By applying this result at a countable dense set of continuity points of F, we obtain the conclusion.

In statistical language, we have proved that F_N is a (strongly) consistent estimator of F as $n \to \infty$, and incidentally that it is unbiased in a pointwise sense. It is then natural to ask how its efficiency compares with that of the estimator obtained from values of the statistic calculated on disjoint subsets of observations. The answer is that the variance is generally of the same order of magnitude, although of course the precise answer to the question is dependent on the context in which it is asked. The

https://doi.org/10.1017/S0305004100049513 Published online by Cambridge University Press

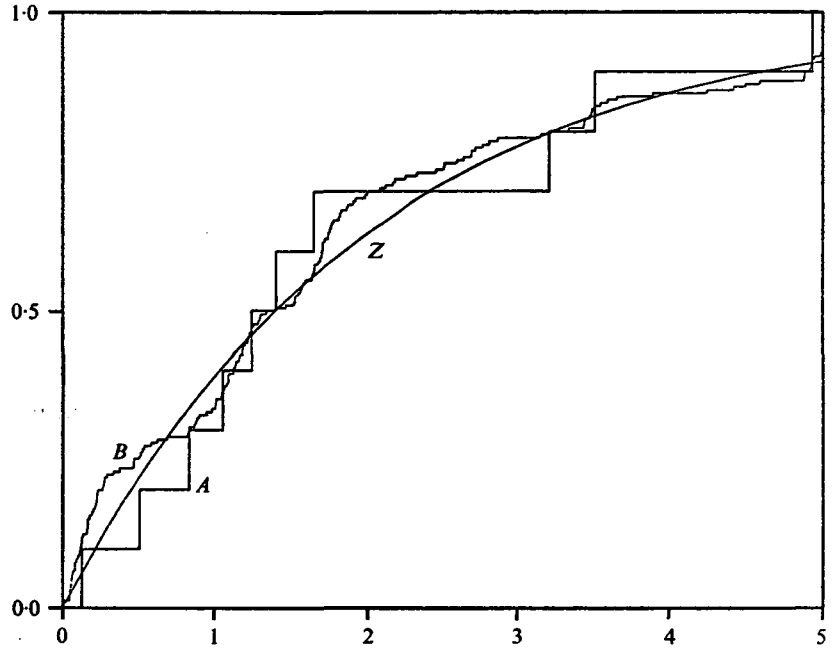


Fig. 1

estimator obtained from overlapping subsets is generally a very much smoother distribution because of the large number of values on which it is based.

4. *Distance distributions.* The example which led us to investigate the above results arises in data analysis, in the area of multidimensional scaling. It is as follows. Suppose we have a fixed probability measure on k-dimensional Euclidean space. If we select a pair of observations independently from this distribution, we can calculate their squared Euclidean distance, and thus we generate a distribution of squared distances. In a few simple cases it is possible to work this out explicitly; if, for example, the parent distribution is spherical normal, then the squared-distance distribution is chi-squared. Now suppose that we select not two but n independent observations. We can calculate $\frac{1}{2}n(n-1)$ different squared distances; is it true that as $n \to \infty$ the EDF of these converges to the distribution of a single squared distance? Our convergence theorem shows that this is indeed the case. Figure 1 illustrates this in a specific case. 20 (pseudo-) random independent $N(0, 1)$ variables $\{y_i: i = 1, \ldots, 20\}$ were generated and used to construct graphs as follows.

Curve A is the EDF of the 10 independent identically distributed χ^2_2 variables $\{a_i = y^2_{2i-1} + y^2_{2i}: i = 1, \ldots, 10\}$.

Curve B is the EDF of the off-diagonal dissociated family

$$\{b_{ij} = y_i^2 + y_j^2: 1 \leqslant i < j \leqslant 20\}.$$

Curve Z is the function $u = 1 - e^{-\frac{1}{2}z}$, which is the DF of χ^2_2.

https://doi.org/10.1017/S0305004100049513 Published online by Cambridge University Press