NOTES

A NOTE ON WEIGHTED RANDOMIZATION[1]

By D. R. Cox[2]

Princeton University

Summary. It is shown that in simple statistical designs in which a covariance adjustment is made for concomitant variation, an unbiased between-treatment mean square can be produced by weighted randomization, i.e., by selecting an arrangement at random from a set of arrangements, giving different arrangements in the set unequal chances of selection.

1. Introduction. Randomization is one of the key elements in the statistical aspects of experimental design [2]. It has as its object the conversion, under rather weak assumptions, of uncontrolled variation of whatever form into effectively random variation. It thus makes the conclusions drawn from the experiment more objective and avoids the introduction of strong, and quite often unrealistic, assumptions about the uncontrolled variation. The methods of randomization in practical use depend on selecting one arrangement from a set giving each arrangement in the set equal chance of selection. For example in a randomized block design, the set would usually be that of all randomized block designs obtained by permuting the treatments within the chosen grouping of units into blocks. An arrangement for use would be one such design selected at random out of the set. The purpose of the present note is to point out the theoretical advantage in certain cases of choosing from the set with unequal probabilities. No recommendation is made about what should be done in practice in such situations.

The following assumption will be made throughout. We have N experimental units (plots, animals, etc.) and r alternative treatments to be compared, one treatment being applied to each unit. Suppose that there is a quantity z_i associated with the ith unit and a constant a_μ associated with the μth treatment, such that if the μth treatment is applied to the ith unit, the resulting observation will be

$$z_i + a_\mu ; \tag{1}$$

independently of the particular allocation of treatments to the other units. The z_i, a_μ are indeterminate to within a constant. The object of the experiment is considered to be the estimation, and possibly significance-testing, of linear contrasts among the a_μ. Assumption (1) can easily be generalized without

Received August 19, 1955.

[1] Research supported in part by the Office of Naval Research.

[2] Present address, Birkbeck College, University of London.

affecting the arguments that follow; for example a completely random term of constant mean and variance can be added to (1), but this will not be done here.

The essential points of (1) are that we are measuring on a scale on which treatment and unit terms are additive, that the treatment effects are constant, and that there is no competition or interference between different units.

The design is called unbiased [6] if it is possible to calculate from the observations the following:

(i) unbiased estimates of the linear contrasts among the a_μ, for example of the differences, $a_\mu - a_\nu$;

(ii) unbiased estimates of the variances of the estimates in (i);

(iii) a mean square between treatments, s_b^2, and the mean square for residual, s_r^2, such that the expectation of s_b^2 is greater than or equal to that of s_r^2, with equality if and only if $a_1 \cdots = a_\tau$.

It is well known that most of the standard designs are unbiased under Assumption (1); the quantities in (i)–(iii) are calculated by the usual analysis of variance methods, and expectations are taken over the set of arrangements from which the one actually used has been selected at random. We shall assume for the purpose of this paper that one requirement for a design to be satisfactory is that it should be unbiased.

Four examples of methods of design that are not unbiased in this sense under simple randomization are

(a) a completely randomized, or other simple design, in which adjustments for a concomitant variable are made by analysis of covariance [5];

(b) the so-called semi-Latin square [6];

(c) a Latin square type cross-over design in which Assumption (1) is extended to allow for simple carry-over of treatment effects from one period to the next [5], [6];

(d) certain designs in which for some practical reason there is a severe restriction on the treatment arrangements that are admissible. An example is [1].

This note is concerned with (a).

2. Adjustment for a concomitant variable. Suppose that on each experimental unit, a concomitant variable is measured, giving a value x_i for the ith unit, and that $x_1, \cdots, x_N$ are fixed and independent of the allocation of treatments to units. Nothing is assumed in the randomization analysis about the relation between the x's and the z's.

Consider to begin with a completely randomized experiment with n units for each treatment, $N = n\tau$. Denote the main observations, given by (1), by $y_1, \cdots, y_N$. These are random variables depending on the particular arrangement of treatments selected. Let $\sum_\mu$ denote summation over those units receiving the μth treatment. In the usual way define

$$(2) \qquad \bar{y}_\mu = \textstyle\sum_\mu y_i / n, \qquad \bar{x}_\mu = \sum_\mu x_i / n,$$
$$\bar{y}_. = \textstyle\sum y_i / N, \qquad \bar{x}_. = \sum x_i / N,$$

and set up the analysis of covariance table

		x^2	xy	y^2	d.f.
	Between treatments	B_{xx}	B_{xy}	B_{yy}	$\tau - 1$
(3)	Residual	R_{xx}	R_{xy}	R_{yy}	$\tau(n - 1)$
	Total	T_{xx}	T_{xy}	T_{yy}	$\tau n - 1$

where, for example,

$$R_{xy} = \sum_{\mu=1}^{\tau} \sum_{\mu} (x_i - \bar{x}_\mu)(y_i - \bar{y}_\mu). \tag{4}$$

Let $b_r = R_{xy} / R_{xx}$ and define the adjusted treatment means

$$\hat{y}_\mu = \bar{y}_\mu - b_r(\bar{x}_\mu - \bar{x}_.). \tag{5}$$

Also define an estimate of the variance of $\hat{y}_\mu - \hat{y}_\nu$ by

$$V_e(\hat{y}_\mu - \hat{y}_\nu) = s_r^2\left\{\frac{2}{n} + \frac{(\bar{x}_\mu - \bar{x}_\nu)^2}{R_{xx}}\right\}, \tag{6}$$

with

$$\operatorname*{Av}_{\mu,\nu} V_e(\hat{y}_\mu - \hat{y}_\nu) = s_r^2\left\{\frac{2}{n} + \frac{2}{n(\tau - 1)} \frac{B_{xx}}{R_{xx}}\right\}, \tag{7}$$

where

$$s_r^2 = \frac{1}{(n\tau - \tau - 1)} (R_{yy} - R_{xy}^2/R_{xx}) \tag{8}$$

is the residual mean square of y adjusting for regression on x. Finally the mean square for treatments adjusting for regression on x is

$$s_b^2 = \frac{1}{(\tau - 1)}\left(T_{yy} - \frac{T_{xy}^2}{T_{xx}} - R_{yy} + \frac{R_{xy}^2}{R_{xx}}\right). \tag{9}$$

The definition of these quantities is based on the least-squares theory of analysis of covariance; we now consider the randomization theory.

3. A method for calculating expectations under randomization. To investigate randomization expectations an elegant and powerful method due to Grundy and Healy [3] will be used. Denote the expectation under simple (unweighted) randomization of any function, f, of the observations by $E_P(f)$,

$$E_P(f) = \frac{1}{(\text{no. of possible arrangements})} \times \sum_{\text{all arrangements}} f. \tag{10}$$

Consider, as an example of the method, the calculation of $E_P(R_{yy})$, when the design is completely randomized. It is easily seen from (1) that this expectation is independent of $\alpha_1, \cdots, \alpha_r$ and is a homogeneous completely symmetric function of $z_1, \cdots, z_N$ of degree two, invariant under translation of the z's.

Therefore

$$E_P(R_{yy}) = \alpha \sum_{i=1}^{N} (z_i - \bar{z}_.)^2, \tag{11}$$

identically in $z_1, \cdots, z_N$, where α is a constant. If $z_1, \cdots, z_N$ is a random sample from a population of variance σ^2, the expectations of the left- and right-hand sides of (11) are respectively $\tau(n - 1)$ and $\alpha(\tau n - 1)$, whence $\alpha = \tau(n - 1) / (\tau n - 1)$. Thus

$$E_P(R_{yy}) = \frac{\tau(n-1)}{\tau n - 1} \sum_{i=1}^{N} (z_i - \bar{z}_.)^2. \tag{12}$$

The choice of $z_1, \cdots, z_N$ as a random sample in the last step of the argument has no physical significance and is purely a mathematical device to exploit knowledge of the behavior of R_{yy} under the usual hypotheses of least-squares theory; see also [4].

In general the method is to establish the general form of the expectation by considerations of symmetry and invariance and then to find the precise expression by special choice of the z's, exploiting our knowledge of what happens under the conditions of least-squares theory. Thus suppose that we require to show that for a Latin square the expectations of the mean squares for treatments and residual are equal, when there are no treatment effects. Consideration of symmetry and invariance show that both expectations are multiples of the residual sum of squares of the z's, considered as a row $\times$ column arrangement. Equality of the two expectations under least-squares theory then proves this equality under general randomization theory.

The result (12) is well known and can be obtained directly without difficulty. The point of Grundy and Healy's method is that it avoids enumerative calculations, and its advantage is consequently greater in the more complicated situations, such as, for example, in the proofs that Latin squares, balanced incomplete blocks, and so on are unbiased under (1).

4. The application to covariance adjustments. If we try to calculate the randomization expectations of s_r^2, s_b^2, defined in (8) and (9), there is the difficulty that R_{xy}^2 / R_{xx} is a ratio of random variables so that no simple exact expression for the form of its expectation can be written down. When $\alpha_1 = \cdots = \alpha_r$, T_{xx}, $T_{xy} = T_{xz}$, $T_{yy} = T_{zz}$ are constant and it follows from (8), (9), and (12) that $E_P(s_r^2) = E_P(s_b^2)$ if and only if $E_P(R_{xy}^2 / R_{xx})$ is linearly related in a particular way to T_{zz} and T_{xz}^2 / T_{xx}. Consideration of the form of $E_P(R_{xy}^2 / R_{xx})$ shows that no such relation can hold identically in the x's and the z's. Hence there is, in general, bias, although we always have

$$E_P(\hat{y}_\mu - \hat{y}_\nu) = a_\mu - a_\nu . \tag{13}$$

The bias arises from the factor $1/R_{xx}$ in (8) and so it is natural to try to remove the bias by weighting each arrangement of treatments proportionally to R_{xx}. This is the general idea behind the following considerations. Suppose that the values $x_1, \cdots, x_N$ are available to the experimenter prior to the allocation of treatments to units. Let w by any non-negative function of $x_1, \cdots, x_N$, defined for each arrangement of treatments within the set in which each treatment occurs n times. Let an arrangement be selected for use giving each design in the set a probability of selection proportional to w; we shall call this a process of weighted randomization using w as weight function. If f is any function of the observations y and x, its expectation under weighted randomization is $E_W(f)$, where

$$E_W(f) = E_P(wf)/E_P(w). \tag{14}$$

Let $w = R_{xx}$ and consider $E_W(s_r^2)$. By (15) we need to know $E_P(R_{xx}R_{yy} - R_{xy}^2)$. This is independent of $a_1, \cdots, a_r$ and is homogeneous and of degree two in $x_1, \cdots, x_N$ and in $z_1, \cdots, z_N$ separately, and is unaffected by interchanging the x's with the z's. Hence

$$\begin{aligned} E_P(R_{xx}R_{yy} - R_{xy}^2) = {} & A\bar{x}_.^2\bar{z}_.^2 \\ & + B(\bar{x}_.^2 T_{zz} + \bar{z}_.^2 T_{xx}) \\ & + CT_{xx}T_{zz} + DT_{xz}^2 \\ & + E\sum_{i=1}^{N}(x_i - \bar{x}_.)^2(z_i - \bar{z}_.)^2 \\ & + F\bar{x}_.\bar{z}_.T_{xz} \\ & G\left\{\bar{x}_.\sum_{i=1}^{N}(x_i - \bar{x}_.)(z_i - \bar{z}_.)^2 + \bar{z}_.\sum_{i=1}^{N}(x_i - \bar{x}_.)^2(\bar{z}_i - \bar{z}_.)\right\}, \end{aligned} \tag{15}$$

where $A, \cdots, G$ are constants. The simplest way of verifying (15) from first principles is to note that

$$\sum x_i^2 z_i^2, \qquad \sum (x_i^2 z_i z_j + x_i x_j z_i^2), \qquad \sum (x_i^2 z_k z_l + x_k x_l z_i^2),$$
$$\sum x_i x_j z_i z_j, \qquad \sum x_i x_j z_i z_k, \qquad \sum x_i x_j z_k z_l$$

are the seven types of sum with the requisite degree of symmetry and that the right-hand side of (15) has seven arbitary constants. $R_{xx}R_{yy} - R_{xy}^2$ is unaffected by changing x_i to $x_i + a$ and z_i to $z_i + b$, $i = 1, \cdots, N$. Since this is true identically in the x's and z's, $A = B = F = G = 0$. Next, if $z_i = \lambda x_i$, $i = 1, \cdots, N$, $R_{xx}R_{yy} - R_{xy}^2$ is identically zero, so that

$$0 = \lambda^2\left\{CT_{xx}^2 + DT_{xx}^2 + E\sum_{i=1}^{N}(x_i - \bar{x}_.)^4\right\}, \tag{16}$$

whence $E = 0$. $C = -D$. If we combine this result with the expression for $E_P(R_{xx})$ corresponding to (12), we have

$$E_W(s_r^2) = H(T_{zz} - T_{xz}^2 / T_{xx}), \tag{17}$$

where H is a constant. Finally let $x_1, \cdots, x_N$ be arbitrary but fixed and let the z_i be uncorrelated random variables with means βx_i and constant variance σ^2. If E denotes expectation over their distribution,

$$E(T_{zz} - T_{xz}^2 / T_{xx}) = (n\tau - 2)\sigma^2,$$
$$E(s_r^2) = \sigma^2,$$

by the ordinary theory of regression. Finally, since $EE_W(s_r^2) = E_W E(s_r^2)$, (17) leads to $H = 1/(n\tau - 2)$, so that

$$E_N(s_r^2) = \frac{1}{(n\tau - 2)}\left(T_{zz} - \frac{T_{xz}^2}{T_{xx}}\right) = \sigma_r^2, \tag{18}$$

say.

Similarly if $a_1 = \cdots = a_\tau$,

$$E_W(s_b^2) = \sigma_r^2, \tag{19}$$

the unbiased property. When the a's are not all equal, a multiple of their corrected sum of squares is added to (19); details will not be given.

We now investigate the corresponding theory for the variance and estimated variance of the difference between two adjusted treatment means, $\hat{y}_\mu - \hat{y}_\nu$, say. It does not seem possible to obtain exact results corresponding to those for s_b^2 and s_r^2 and we shall need to use the following asymptotic results. If, as N tends to infinity, f and g are random variables, functions of the y's and the x's with fixed means and with variance of order $1/N$, then

$$E(fg) = E(f)E(g) + O(1/N), \tag{20}$$

and, under weak conditions on g,

$$E(f/g) = E(f) / E(g) + O(1/N). \tag{21}$$

These will be used with E standing for E_P or E_W as convenient. The expectation on the left of (21) is to be taken as referring to the asymptotic distribution of f/g.

Now we have from (7), (18), and (20) that

$$E_W[\underset{\mu,\nu}{\text{Av}}\, V_e(\hat{y}_\mu - \hat{y}_\nu)] = \frac{2\sigma_r^2}{n} + \frac{2\sigma_r^2}{n(\tau - 1)} E_W\left(\frac{B_{xx}}{R_{xx}}\right)\left(1 + O\left(\frac{1}{N}\right)\right) \tag{22}$$

$$= \frac{2\sigma_r^2}{n}\left\{1 + \frac{1}{\tau(n-1)}\left(1 + O\left(\frac{1}{N}\right)\right)\right\}. \tag{23}$$

If τ is fixed and n tends to infinity, the relative error in (20) is of order $1/N^2$. In obtaining (22) we have assumed that the x's and the z's are such as to make the variances of s_r^2 and B_{xx} / R_{xx} of order $1/N$.

Similarly to find the actual variance of $\hat{y}_\mu - \hat{y}_\nu$ we have that

$$\begin{aligned}\underset{\mu,\nu}{\text{Av}}\, \{(\hat{y}_\mu - \hat{y}_\nu) - (a_\mu - a_\nu)\}^2 &= \underset{\mu,\nu}{\text{Av}}\left\{\frac{1}{n}(\Sigma_\mu z - \Sigma_\nu z) - \frac{b_r}{n}(\Sigma_\mu x - \Sigma_\nu x)\right\}^2 \\ &= \frac{2}{n(\tau - 1)}\left(B_{zz} - \frac{2R_{xz}B_{xz}}{R_{xx}} + \frac{R_{xz}^2 B_{xx}}{R_{xx}^2}\right).\end{aligned} \tag{24}$$

If we apply the operator E_W to (24), we get the required variance. The expectation can be evaluated in a way similar to (23), dealing with the last term by (21). The final answer is the right-hand side of (23). That is, to the order indicated above, (7) is an unbiased estimate of the average variance of $\hat{y}_\mu - \hat{y}_\nu$.

5. Extensions. The calculations in Section 4 have, for simplicity, been made for the completely randomized design. However the results can be extended to designs such as randomized blocks and Latin squares; weighting proportional to the residual sum of squares of x again gives an unbiased treatment mean square. Another generalization is to multiple analysis of covariance, in which the treatment means are adjusted for k concomitant variables $x_1, \cdots, x_k$. The appropriate weighting function is then the residual generalized variance, i.e., the determinant $|R_{ij}|$, where R_{ij} is the residual sum of products of x_i and x_j.

6. Discussion. The idea of weighted randomization discussed above is probably solely of theoretical interest, at any rate in the context considered here. A full discussion of possible practical applications would require further work, but the following points are worth making.

(i) The bias in unweighted randomization is probably small, except possibly when N is very small and the correlation between the z's and the x's very non-linear. Further work is needed, however, to find the likely magnitude of the bias in typical cases.

(ii) Weighted randomization is perhaps most likely to be of practical value when a series of similar experiments are planned, each with a small value of N. Another possible application is to Latin square designs in which it is desired to control variations diagonally across the square, in addition to row and column variation. This can be done by inserting a suitable concomitant variable, for example the product of row number and column number suitably coded. Weighted randomization would justify such a method in the same way that ordinary randomization justifies the conventional use of the Latin square.

(iii) Arrangements with a large value of R_{xx} will have a small value for B_{xx} and conversely. Hence the weighting proportional to R_{xx} attaches greater chance of selection to those arrangements in which the treatment groups are balanced with respect to the mean value of x.

(iv) If weighted randomization is to be done in practice with N not very small, some short-cut method is needed for selecting an arrangement, since the enumeration of all arrangements and the calculation of R_{xx} for each would usually be too tedious. Professor J. W. Tukey has pointed out that weighted randomization can be done reasonably simply as follows. Let M be the maximum over-all arrangement of R_{xx}. Select an arrangement by unweighted randomization and calculate R_{xx} for it. Reject the arrangement with probability $1 - R_{xx} / M$. Continue until an arrangement is accepted.

(v) Weighted randomization is, of course, restricted to cases in which the concomitant variable is available prior to the allocation of treatments to units.

REFERENCES

[1] D. R. Cox, "The design of an experiment in which certain treatment arrangements are inadmissible," *Biometrika*, Vol. 41 (1954), pp. 287–295.

[2] R. A. Fisher, *The Design of Experiments*, 5th ed., Oliver and Boyd, Ltd., 1949.

[3] P. M. Grundy and M. J. R. Healy, "Restricted randomization and quasi-Latin squares," *J. Roy. Stat. Soc.* (B), Vol. 12 (1950), pp. 286–291.

[4] J. O Irwin and M. G. Kendall, "Sampling moments of moments for a finite population," *Ann. Eugenics* Vol. 12 (1943–45), pp. 138–142.

[5] O. Kempthorne, *The Design and Analysis of Experiments*, J. Wiley and Sons, 1952.

[6] F. Yates, "Bases logiques de la planification des expériences, *Ann. Inst. H. Poincaré* Vol. 12 (1951), pp. 97–112.

ON STOCHASTIC APPROXIMATION METHODS[1]

By J. Wolfowitz

Cornell University

In [1] A. Dvoretzky proved the theorem quoted below, which implies all previous results on the convergence to a limit of stochastic approximation methods. (For a description of these results see [1].) In the present note we give a simple and, we think, perspicuous proof of this theorem which may be of help in further work. The present note is entirely self-contained and may be read without reference to [1].

Theorem. (Dvoretzky) *Let* α_n, β_n *and* $\gamma_n (n = 1, 2, \cdots)$ *be non-negative real numbers satisfying*

$$\lim_{n=\infty} \alpha_n = 0, \tag{1}$$

$$\sum_{n=1}^{\infty} \beta_n < \infty, \tag{2}$$

and

$$\sum_{n=1}^{\infty} \gamma_n = \infty. \tag{3}$$

Let θ *be a real number and* $T_n (n = 1, 2, \cdots)$ *be measurable transformations satisfying*

$$|T_n(r_1, \cdots, r_n) - \theta| \leqq \max[\alpha_n, (1 + \beta_n)|r_n - \theta| - \gamma_n] \tag{4}$$

for all real $r_1, \cdots, r_n$. *Let* X_1 *and* $Y_n (n = 1, 2, \cdots)$ *be random variables and define*[2]

Received December 16, 1955.

[1] This research was supported by the United States Air Force under Contract No. AF18-(600)-685 monitored by the Office of Scientific Research.

[2] In the proof of the theorem we will, for the sake of brevity, write $T_n(X_n)$ for

$$T_n(X_1, \cdots, X_n),$$

just as is done in [1]. No ambiguity will be caused by this.