Resampling and exchangeable arrays

PETER MCCULLAGH

Department of Statistics, University of Chicago, 5734 University Avenue, Chicago IL 60637, USA. E-mail: pmcc@galton.uchicago.edu

The nonparametric, or resampling, bootstrap for a single unstructured sample corresponds to the algebraic operation of monoid composition, with a uniform distribution on the monoid. With this interpretation, the notion of resampling can be extended to designs having a certain group-invariance property. Two types of exchangeable array structures are considered in some detail, namely the one-way layout, and the two-way row–column exchangeable design. Although in both cases there is a unique group under which the sampling distribution of the observations is exchangeable, the choice of monoid is not unique. Different choices of monoid can lead to drastically different, and in some cases quite misleading, inferences.

Keywords: bootstrap; exchangeability; group invariance; monoid composition; product monoid; symmetric function

1. Introduction

Suppose that inference is required for a scalar parameter θ in the form of confidence intervals. A consistent, asymptotically normal, estimator T is given, but no consistent estimate of variance is available. Under what conditions can resampling techniques be used to approximate the distribution of T, and thereby generate a first-order correct confidence interval?

Two circumstances in which problems of this nature can arise are as follows:

 (i) The distribution of the data has a simple structure, possibly with independent components, but the functional form of T is complicated.
 (ii) The functional form of T is simple, possibly even linear, but the dependence structure in the data is sufficiently complicated that it is not feasible to derive a consistent variance estimate.

Most resampling investigations have focused on problems of the first type. The present paper aims at rather specialized problems of the second type in which the dependence among observations is specified by group invariance. That is to say, we consider structures in which Y has the same distribution as gY for each g in the given group $\mathscr{G}$. We do this by first interpreting the standard resampling bootstrap as monoid composition in which the bootstrap distribution is uniform on the monoid. In this context, a monoid is any semigroup that includes $\mathscr{G}$. The purpose, as always, is to generate pseudo-observations in such a way that the bootstrap distribution of a statistic is a faithful estimate of the sampling distribution of the same statistic (Efron and Tibshirani 1993, Chapter 2).

Two specific cases are considered in detail. For the one-way layout, we consider two linear statistics and six monoids, all containing the relevant wreath product group. All six bootstrap variances are calculated analytically for both statistics. Unfortunately, it appears that the choice of monoid is critical, and what appears to be the most natural choice does not necessarily give the best variance estimate.

For the two-way row–column exchangeable array, one statistic and two monoids are considered. Although there exists a satisfactory consistent variance estimate, neither bootstrap scheme produces a consistent estimate. In fact, it is shown that no resampling scheme, of the type considered in this paper, can yield a consistent estimate of $\mathrm{var}(\overline{Y})$ in row–column exchangeable arrays.

Finally, we conclude with some remarks concerning exchangeability in regression models.

2. Resampling

2.1. Monoid composition

Let Ω be the set of n observed subjects or sampling units. The observation vector y is a real-valued function on Ω, a point in the vector space $\mathscr{R}^{\Omega}$. The value of y on unit ω is usually denoted by $y(\omega) = y_{\omega}$, and the list of these values is a vector in $\mathscr{R}^{\Omega}$.

Let $\mathscr{G}$ be the symmetric group acting on Ω: each element in $\mathscr{G}$ is an invertible function from Ω to itself. The composition of y with $\varphi \in \mathscr{G}$ is a new function $y \circ \varphi$ from Ω to $\mathscr{R}$, whose value on unit ω is given by

$$(y \circ \varphi)(\omega) = y_{\varphi(\omega)}.$$

Evidently, $y \circ \varphi$ is a permutation of y. Further, if φ is chosen uniformly at random from the group, $y \circ \varphi$ is a random permutation. For given y, the distribution of the statistic $T(y \circ \varphi)$ is called the permutation distribution of T. Evidently the permutation distribution is discrete, with at most $n!$ points of support.

Let $\mathscr{M}$ be the set of all n^n transformations of Ω to itself. This set includes the group, but it also includes the non-invertible transformations in which several points in Ω are mapped to the same image. The composition $y \circ \varphi$ is a new function from Ω to $\mathscr{R}$. For statistical purposes, $y \circ \varphi$ is a sample of the y values taken *with replacement*. In particular, if φ is chosen uniformly at random from $\mathscr{M}$, $y \circ \varphi$, considered as a function from $\mathscr{M}$ into $\mathscr{R}^{\Omega}$, is a random variable whose distribution is called the bootstrap distribution of y.

In algebraic terminology, $\mathscr{M}$ is a monoid (semigroup with identity) acting on Ω. There are in fact many monoids acting on Ω: the symmetric group itself is an example of a monoid. The particular monoid described above, the one that occurs in bootstrap resampling, is called the full monoid on Ω.

2.2. Bootstrap exchangeability

A list of n independent and identically distributed (i.i.d.) random variables indexed by Ω is a

vector in $\mathscr{R}^\Omega$ whose sampling distribution is exchangeable, or invariant under permutation. In other words, for each g in the symmetric group, Y and $Y \circ g$ have the same sampling distribution. Under certain conditions, this property of the sampling distribution is transferred to permutation and bootstrap distributions. For fixed y, the bootstrap random variable is a function

$$Y^*(\varphi) = y \circ \varphi$$

from the monoid $\mathscr{M}$ into $\mathscr{R}^\Omega$. The composition of Y^* with a group element g is given by

$$Y^* \circ g = (y \circ \varphi) \circ g = y \circ (\varphi \circ g) = y \circ \varphi'.$$

If, for each $\varphi \in \mathscr{M}$, all elements in the set

$$\{\varphi \circ g : g \in \mathscr{G}\},$$

have equal probability, then the bootstrap distribution is exchangeable. In general, the distribution need not be uniform on $\mathscr{M}$, but it must be uniform on each of the group orbits in $\mathscr{M}$.

To see what the group orbits of $\mathscr{M}$ are, consider an example in which Ω is a set of four integers. Consider the functions

$$\varphi_1 = \{1 \mapsto 4, 2 \mapsto 4, 3 \mapsto 3, 4 \mapsto 2\},$$

$$\varphi_2 = \{1 \mapsto 1, 2 \mapsto 2, 3 \mapsto 3, 4 \mapsto 3\}.$$

The inverse functions are

$$\varphi_1^{-1} = \{1 \mapsto \varnothing, 2 \mapsto \{4\}, 3 \mapsto \{3\}, 4 \mapsto \{1, 2\}\},$$

$$\varphi_2^{-1} = \{1 \mapsto \{1\}, 2 \mapsto \{2\}, 3 \mapsto \{3, 4\}, 4 \mapsto \varnothing\},$$

so $\varphi_1^{-1}(\Omega)$ is the partition $12|3|4$, and $\varphi_2^{-1}(\Omega)$ is the partition $1|2|34$. These set partitions have the same block sizes corresponding to the number partition $4 = 2 + 1 + 1$, so they belong to the same group orbit.

The group itself constitutes one such orbit, so the permutation distribution is exchangeable. The number of orbits, or points in $\mathscr{M}/\mathscr{G}$, is equal to the number of partitions of the number n. The group itself corresponds to the orbit (1^n) in which each element of Ω has a distinct image. The orbit $(2, 1^{n-2})$ is that subset of $\mathscr{M}$ in which two points have the same image and all other points have distinct images. The bootstrap distribution is exchangeable if and only if the distribution on $\mathscr{M}$ is uniform on each of these $p(n)$ orbits. The uniform distribution on $\mathscr{M}$ satisfies the conditions for exchangeability, so the bootstrap distribution is exchangeable.

2.3. Symmetric functions and bootstrap cumulants

The bootstrap distribution is the uniform distribution on the monoid. Another way of saying the same thing is to express the bootstrap resampling scheme componentwise as

$$y^{*i} = \varphi_t^i y^t,$$

with implicit summation over repeated indices. Independently for each i, φ^i is a random variable whose distribution is uniform multinomial on Ω. The joint cumulants of degree r are

$$\text{cum}_r(\varphi_{t_1}^{i_1}, \ldots, \varphi_{t_r}^{i_r}) = \delta^{i_1 \ldots i_r} \Gamma_{t_1, \ldots, t_r},$$

where δ is the indicator function for equality of the indices, and the Γs are the cumulants of the uniform multinomial. The multinomial moments are $\Gamma_{t_1, \ldots, t_r} = \delta_{t_1 \ldots t_r}/n$, from which we obtain

$$\Gamma_r = \delta_r/n$$

$$\Gamma_{r,s} = \delta_{rs}/n - \delta_r \delta_s/n^2$$

$$\Gamma_{r,s,t} = \delta_{rst}/n - \delta_r \delta_{st}[3]/n^2 - 2\delta_r \delta_s \delta_t/n^3$$

and so on, using the formulae for cumulants in terms of moments.

Let k_r be the sample k-statistic of degree r, so k_1 is the sample mean and k_2 the sample variance. In general, k_r is the unique symmetric polynomial of degree r whose i.i.d. sampling expectation is κ_r, the rth cumulant of the sampling distribution. Denote by k_r^* the random variable $k_r(y^*)$. A straightforward calculation shows that the bootstrap distribution satisfies the following identities:

$$\text{E}(k_1^*|y) = k_1,$$

$$n \operatorname{var}(k_1^*|y) = \text{E}(k_2^*|y) = (n-1)k_2/n,$$

$$n^2 \operatorname{cum}_3(k_1^*|y) = n \operatorname{cov}(k_1^*, k_2^*|y) = \text{E}(k_3^*|y) = (n-1)(n-2)k_3/n^2,$$

$$n^3 \operatorname{cum}_4(k_1^*|y) = n^2 \operatorname{cum}_3(k_1^*, k_1^*, k_2^*|y) = n \operatorname{cov}(k_1^*, k_3^*|y)$$

$$= n \operatorname{var}(k_2^*|y) = \text{E}(k_4^*|y) = (n-1)((n^2 - 6n + 6)k_4 - 6nk_{22})/\text{n}^3.$$

In the final expression, k_{22} is the unique symmetric function (polykay) whose i.i.d. expectation is κ_2^2 (McCullagh 1987, Chapter 4). More generally, for any linear statistic $T^* = l_i y^{*i}$, we have

$$E(T^*|y) = \sum_i l_i E(k_1^*|y),$$

$$n \operatorname{var}(T^*|y) = \sum_i l_i^2 E(k_2^*|y),$$

$$n^2 \operatorname{cum}_3(T^*|y) = \sum_i l_i^3 E(k_3^*|y),$$

$$n^3 \operatorname{cum}_4(T^*|y) = \sum_i l_i^4 E(k_4^*|y),$$

and so on for all higher-order cumulants. As we shall see in Section 5, these are special cases of identities satisfied by some, but not all, bootstrap resampling schemes.

2.4. Bootstrap

In the sections that follow, we consider random arrays whose sampling distribution is invariant under a specified group, $\mathcal{G}$. In this paper, a bootstrap distribution is defined as follows. First, we choose a monoid $\mathcal{M}$ such that $\mathcal{G} \subset \mathcal{M}$. Second, we choose a probability distribution on the monoid, uniform on group orbits and independent of y. Given y, the bootstrap distribution is the distribution of $y \circ \varphi$, where φ has the specified distribution on $\mathcal{M}$. In most cases, the distribution is uniform on the monoid, but this is not critical.

3. The one-way exchangeable array

3.1. Resampling schemes

Consider an array of b blocks, each of size n, having the property that the permuted array of random variables has the same joint sampling distribution as the original array. In this context, permutation refers to the wreath product group in which blocks are permuted by an element φ from S_b, the symmetric group acting on the blocks. Observations in each block are permuted independently by composition with $\psi \in S_n$. In other words, each permutation π has components $(\varphi, \psi_1, \ldots, \psi_b)$ where φ permutes blocks, and ψ_i permutes observations within a block. The action on the array is given either by

$$\pi : (i, j) \mapsto (\varphi(i), \psi_i(j)) \tag{1}$$

or by

$$\pi : (i, j) \mapsto (\varphi(i), \psi_{\varphi(i)}(j)), \tag{2}$$

depending on the order of operations. In other words, under (1), the (i, j)th element of $y \circ \pi$ is $y_{\varphi(i), \psi_i(j)}$: under (2) we get $y_{\varphi(i), \psi_{\varphi(i)}(j)}$, which is different. The sampling distribution of Y is exchangeable if, for each π in the group, the distribution of $Y \circ \pi$ is the same as the distribution of Y. So far as the definition of exchangeability is concerned, the actions (1) and (2) are effectively equivalent because to each π in the group, there corresponds a π', given by

$$\pi' = (\varphi, \psi_{\varphi(1)}, \ldots, \psi_{\varphi(b)}),$$

such that $y \circ \pi'$ using (1) is equal to $y \circ \pi$ using (2). Nevertheless, for semigroup operations, it is essential to distinguish between the two modes of action.

There are various ways in which resampling might operate in this context, but all such schemes are based on an extension of the group to a monoid. Six extensions are described and studied:

Boot-I. First permute the observations independently within the blocks. Then pick a random sample of the blocks with replacement.

Boot-II. First pick a random sample of the blocks with replacement. Then permute the observations independently within the blocks.

Boot-III. First permute the blocks. Then pick a random sample with replacement from each of the blocks.

Boot-IV. First select a random sample with replacement from each block. Then pick a random sample of the blocks.

Boot-V. First pick a random sample of the blocks with replacement. Then pick a random sample with replacement from each of the blocks.

Boot-VI. Disregard the block structure, select a random sample of size nb with replacement, and arrange in blocks of size n.

Each of these resampling schemes consists of two parts. The first part is a set of transformations from the set of labels $\{(i, j)\}$ to itself. The second part is a distribution on this set of transformations. Each set of transformations is a monoid containing the wreath product group as a subset. Each bootstrap distribution is uniform on the associated monoid, and hence invariant under the group.

For monoids II, III and V, the transformation is given by (1); for I and IV, the transformation is given by (2). In the case of transformation (1), duplicate blocks are permuted or sampled independently. Under (2), however, identical samples are obtained from duplicate blocks. The order of operations is immaterial in the wreath product of groups, but order does matter for semigroups.

Composite resampling schemes can also be constructed in which a randomization device is used to select a resampling scheme from the list above. A die is cast, generating a sequence of independent values. Depending on the outcome of the die, one of the sampling schemes listed above is chosen to generate a transformation φ. This is equivalent to using the full monoid $\mathscr{M}_{VI}$, with a non-uniform distribution on the monoid. The distribution is, however, uniform on each group orbit.

The relations among the monoids are as follows:

$$\mathscr{G} \subset \mathscr{M}_I \subset \mathscr{M}_{II} \subset \mathscr{M}_V \subset M_{VI}, \qquad \mathscr{G} \subset \mathscr{M}_{III} \subset \mathscr{M}_{IV} \subset \mathscr{M}_V \subset M_{VI}.$$

The number of elements in each monoid is as follows:

$$|\mathscr{G}| = b!(n!)^b; \qquad |\mathscr{M}_I| = \sum_{1 \leqslant j \leqslant b} S_{bj} j!(n!)^j; \qquad |\mathscr{M}_{II}| = b^b (n!)^b;$$

$$|\mathscr{M}_{III}| = b! n^{nb}; \qquad |\mathscr{M}_{IV}| = \sum_{1 \leqslant j \leqslant b} S_{bj} j! n^{nj}; \qquad |\mathscr{M}_V| = b^b n^{nb}; \qquad |\mathscr{M}_{VI}| = (nb)^{nb};$$

where S_{nm} is Stirling's number of the second kind.

3.2. Bootstrap distributions

We consider here two scalar statistics, one invariant under the group, the other non-invariant. The sample mean is an invariant statistic that is linear in the observations. For a non-invariant statistic, we suppose that $m_1 + m_2 = n$, and that in each block the first m_1 observations have treatment level 1, and the remainder have treatment level 2. Then, if $\bar{Y}_1$ is the average of the observations on treatment 1, and $\bar{Y}_2$ the average on treatment 2, the treatment difference, $\bar{Y} - \bar{Y}_2$, is a within-blocks contrast that is linear in the data. We distinguish between the

sampling distributions of these statistics, which depend on unknown parameters, and the bootstrap distributions, which are computable but depend on the particular choice of monoid.

To keep the calculations as simple as possible, we focus only on the first two moments. On the assumption that the sampling distribution is exchangeable and that observations in different blocks are independent, the mean and variance of the sample mean are

$$E(\bar{Y}) = \mu, \qquad \text{var}(\bar{Y}) = (n\sigma_r^2 + \sigma_\varepsilon^2)/(nb).$$

Similar calculations show that the mean and variance of $\bar{Y}_1 - \bar{Y}_2$ are

$$\text{E}(\bar{Y}_1 - \bar{Y}_2) = 0, \qquad \text{var}(\bar{Y}_1 - \bar{Y}_2) = \sigma_\varepsilon^2(1/m_1 + 1/m_2)/b.$$

The variance components here are defined by

$$\text{var}(Y_{ij}) = \sigma_r^2 + \sigma_\varepsilon^2, \qquad \text{cov}(Y_{ij}, Y_{ij'}) = \sigma_r^2,$$

for $j \neq j'$.

Both variances can be estimated in very natural ways using the two invariant quadratic forms

$$S_r^2 = \sum y_{i.}^2/n - y_{..}^2/(nb),$$

$$S_e^2 = \sum_{ij} y_{ij}^2 - \sum_i y_{i.}^2/n,$$

whose expectations are

$$\text{E}(S_r^2) = (b-1)(n\sigma_r^2 + \sigma_\epsilon^2),$$

$$\text{E}(S_e^2) = b(n-1)\sigma_\epsilon^2.$$

It is conventional to estimate $\text{var}(\bar{Y})$ by $S_r^2/(nb(b-1))$, and $\text{var}(\bar{Y}_1 - \bar{Y}_2)$ by $S_e^2(1/m_1 + 1/m_2)/(b^2(n-1))$. Within the class of unbiased symmetric quadratic forms, these estimates are unique.

In all cases, the bootstrap mean of $\bar{Y}^*$ is equal to the sample mean, and the bootstrap mean of $\bar{Y}_1^* - \bar{Y}_2^*$ is zero. The bootstrap variances are quadratic functions of the data given in Table 1, where $s^2 = S_e^2/b(n-1)$ is the within-blocks mean square.

For the invariant statistic, monoids I and II provide the best variance estimate, differing from the unbiased estimate by a factor of $(b-1)/b$. Monoids III and VI give misleadingly small variances, even asymptotically. In a certain limiting sense, if $\sigma_r^2 \neq 0$, it can be argued that monoids IV and V give consistent variance estimates, but these are too large in finite samples, and certainly inferior to I and II.

For the non-invariant statistic, the permutation variance and the monoid-II variance are arguably the only correct values. Monoids I, IV and VI produce variances that are misleadingly large.

Fisher's (1935, Section 21) analysis of Darwin's data, based on a randomization argument, is equivalent to resampling from the wreath product group.

Although monoid II gives a bootstrap variance that is a reasonably good approximation to

Table 1. Bootstrap variances of two statistics

Monoid	$B_2(\bar{Y}^*) \times n^2 b^2$	$B_2(\bar{Y}_1^* - \bar{Y}_2^*) \times bm_1 m_2/n$
$\mathscr{G}$	0	s^2
I	nS_r^2	$(2 - b^{-1})s^2$
II	nS_r^2	s^2
III	S_e^2	$s^2(n-1)/n$
IV	$(2 - b^{-1})S_e^2 + nS_r^2$	$(2 - b^{-1})s^2(n-1)/n$
V	$S_e^2 + nS_r^2$	$s^2(n-1)/n$
VI	$S_e^2 + S_r^2$	$(S_e^2 + S_r^2)/(nb)$
unbiased	$nbS_r^2/(b-1)$	s^2

the sampling variance in both cases, it does not appear that there is a uniformly best scheme that is best for all statistics. Further, what appear at first sight to be natural extensions of the group to a monoid (IV and V), can yield quite different bootstrap distributions. In more complicated designs where it is not feasible to compute a table of bootstrap variances and to compare these with the sampling variance of the statistic, it would not be easy to tell which resampling scheme, if any, is appropriate as a reference distribution.

4. Row–column exchangeable arrays

4.1. Introduction

A random $m \times n$ array Y with elements Y_{ij} is said to be row–column exchangeable if, for any permutations $\pi \in S_m$ and $\pi' \in S_n$, the distribution of the permuted array

$$Y \circ \pi = \{Y_{\pi(i)\pi'(j)}\}$$

is the same as the distribution of Y. We assume that the moments of Y exist, at least up to second order, and that inference is required for the mean value $\mu = \mathrm{E}(Y_{ij})$. Exchangeability implies that all components have the same marginal distribution, and hence the same mean. In addition, all second moments are determined by the four parameters

$$\sigma^2 = \mathrm{var}(Y_{ij})$$

$$\sigma_r^2 = \mathrm{cov}(Y_{ij}, Y_{i,j'}), \qquad j \neq j'$$

$$\sigma_c^2 = \mathrm{cov}(Y_{ij}, Y_{i',j}), \qquad i \neq i'$$

$$\rho = \mathrm{cov}(Y_{ij}, Y_{i',j'}), \qquad i \neq i', j \neq j'.$$

Exchangeability does not imply $\rho = 0$, but the applications that we have in mind concern dissociated arrays in which components that are neither in the same row nor in the same

column are independent. Henceforth, therefore, we assume that $\rho = 0$. In addition, we assume positive definiteness for all m, n, which implies that $\sigma_r^2 \geqslant 0$, $\sigma_c^2 \geqslant 0$, and

$$\sigma_\epsilon^2 = \sigma^2 - \sigma_r^2 - \sigma_c^2 \geqslant 0.$$

4.2. Variances and symmetric quadratic forms

The only summary statistics of interest are those that are invariant under the group $S_m \times S_n$ acting on the rows and columns, the so-called symmetric functions. The obvious estimate of μ, the average of the components, $\hat{\mu} = \bar{Y}..$, is a symmetric function of degree 1. Modulo scalar multiples, this is the unique symmetric function of degree 1. Under mild conditions, if m, n are both large, $\hat{\mu}$ is approximately normally distributed. The sampling variance is

$$\text{var}(\hat{\mu}) = (n\sigma_r^2 + m\sigma_c^2 + \sigma_\epsilon^2)/(mn). \tag{3}$$

An estimate of this combination is required in order to set approximate confidence limits for μ.

The space of symmetric quadratic forms has dimension 4. It is spanned by the familiar sums of squares whose expectations under the model are as follows.

Statistic	Expectation
$S_m^2 = Y_{..}^2/(mn)$	$mn\mu^2 + \sigma_\epsilon^2 + n\sigma_r^2 + m\sigma_c^2$
$S_r^2 - \sum Y_{i.}^2/n - S_m^2$	$(m-1)\sigma_\epsilon^2 + n(m-1)\sigma_r^2$
$S_c^2 = \sum Y_{.j}^2/m - S_m^2$	$(n-1)\sigma_\epsilon^2 + m(n-1)\sigma_c^2$
$S_e^2 = \sum Y_{ij}^2 - S_c^2 - S_r^2 - S_m^2$	$(m-1)(n-1)\sigma_\epsilon^2$

It follows that there is a unique symmetric quadratic form whose expectation is the desired combination (3). This quadratic form is a linear combination of mean squares given by

$$(MS_r + MS_c - MS_e)/(mn), \tag{4}$$

Unfortunately, this variance estimate is not guaranteed to be non-negative. The conventional estimate of variance,

$$s^2 = ((MS_r - MS_e)^+ + (MS_c - MS_e)^+ + MS_e)/(mn),$$

is positive, but not quadratic. Nevertheless, if m, n are both large, the probability that s^2 differs from (4) is exponentially small. The interval $\hat{\mu} \pm k_{\alpha/2}s$, contains μ with probability $\alpha + o(1)$. This is a first-order approximate confidence interval for μ.

4.3. Resampling schemes

Let Ω_r be the set of m row labels, Ω_c the set of n column labels, and $\Omega_r \times \Omega_c$ the set of ordered pairs. The full monoids acting on these sets are denoted by $\mathscr{M}_r$, $\mathscr{M}_c$ and $\mathscr{M}_{rc}$, respectively.

Two specific resampling schemes are considered, as follows:

Boot-I. Select a random sample of size mn with replacement from the observed table. In other words, choose φ uniformly at random from $\mathscr{M}_{rc}$.

Boot-II. Select a random sample of the rows, $\varphi_r \in \mathscr{M}_r$, and a random sample $\varphi_c \in \mathscr{M}_c$, of the columns. The bootstrap table is the set of cells (φ_r, φ_c) applied to $\Omega_r \times \Omega_c$. In other words, choose φ uniformly at random from the product monoid $\mathscr{M}_r \times \mathscr{M}_c$ and set $Y^* = y \circ \varphi$.

For given y, the bootstrap distribution of Y^*, or of any derived statistic such as $\bar{Y}^*$, obviously depends on the distribution of φ in $\mathscr{M}_{rc}$. In the first scheme above, the distribution is uniform on the $(mn)^{mn}$ points in $\mathscr{M}_{rc}$: in the second scheme, the distribution is uniform on the $m^m n^n$ points in the subset $\mathscr{M}_r \times \mathscr{M}_c$. In both schemes, the monoid includes the product group, so both bootstrap distributions are row–column exchangeable. In fact, $\mathscr{M}_{rc}$ includes the full symmetric group, so the first bootstrap distribution is completely exchangeable. This property seems undesirable for structured arrays.

The bootstrap mean of $\bar{Y}^*$ is equal to the sample mean in both cases. The bootstrap variances and their expectations are given below.

It can be seen that the Boot-I variance is misleadingly small unless $\sigma_r^2 = \sigma_c^2 = 0$. The

	Variance estimate $\times$ mn	Sampling expectation
Unbiased	$\dfrac{S_r^2}{m-1} + \dfrac{S_c^2}{n-1} - \dfrac{S_e^2}{(m-1)(n-1)}$	$n\sigma_r^2 + m\sigma_c^2 + \sigma_\varepsilon^2$
Boot-I	$S_r^2/mn + S_c^2/mn + S_e^2/mn$	$\sigma_r^2 \dfrac{m-1}{m} + \sigma_c^2 \dfrac{n-1}{n} + \sigma_\varepsilon^2 \dfrac{mn-1}{mn}$
Boot-II	$S_r^2/m + S_c^2/n + S_e^2/mn$	$\sigma_r^2 \dfrac{n(m-1)}{m} + \sigma_c^2 \dfrac{m(n)}{n} + \sigma_\varepsilon^2 \dfrac{3mn - 2m - 2n + 1}{mn}$

Boot-II variance behaves asymptotically as $(n\sigma_r^2 + m\sigma_c^2 + 3\sigma_\epsilon^2)(1 + o_p(1))$. In a certain asymptotic sense, the estimate is consistent except at the point $\sigma_r^2 = \sigma_c^2 = 0$, when the variance is too large by a factor of 3. So, although the Boot-II variance comes close for most parameter values, neither variance estimate is consistent for $\text{var}(\hat{u})$. If the variance components are small, say $\sigma_r^2 = O(n^{-1})$, $\sigma_c^2 = O(m^{-1})$, and $\sigma_\epsilon^2 = O(1)$, neither bootstrap estimator is consistent, but the unbiased estimator and s^2 are quite satisfactory.

4.4. Positive quadratic functions

All of the resampling schemes considered here are generated by monoid composition, $Y^* = y \circ \varphi$, in which the distribution of φ in $\mathscr{M}$ does not depend on y. Given φ, the

transformation from y to Y^* is linear. Since, by assumption, the distribution of φ in $\mathscr{M}$ is independent of y, each bootstrap moment or cumulant of order r is a homogeneous polynomial in y of degree r. In particular, the bootstrap mean of any linear statistic is a linear function of y, and the bootstrap variance of any linear statistic is a positive quadratic function of y. This property holds for all six resampling schemes described in Section 3 as well as Boot-I and Boot-II in Section 4.3.

In the case of the row–column exchangeable array, however, it was observed that there is a unique unbiased symmetric quadratic statistic (4) that estimates the variance of $\hat{\mu}$. However, this estimator is not positive definite, so there is no bootstrap distribution for which this is the variance of $\bar{Y}^*$. In addition, since there are only four invariant quadratic statistics, it is not difficult to show that, within the class of symmetric quadratic forms, every consistent estimator permits negative values. In other words, there does not exist a symmetric positive definite quadratic estimator that is consistent for $n\sigma_r^2 + m\sigma_c^2 + \sigma_\epsilon^2$. In fact, we can dispense with the symmetry condition. If there exists an asymmetric positive definite quadratic estimator, this can be symmetrized by averaging over permutations to yield a symmetric statistic that remains positive definite, quadratic, and also consistent. This is a contradiction, so we conclude that no consistent estimator exists within the class of positive quadratic functions. But every bootstrap variance of $\bar{Y}^*$ is a positive quadratic function of y. So we conclude that no resampling scheme can yield a consistent estimate of the variance of $\hat{\mu}$. It is not simply the case that the two resampling schemes exhibited above give inconsistent variance estimates; all resampling schemes having the property that the distribution of φ is independent of y have the same property.

4.5. Hybrid resampling schemes

Another bootstrap scheme that has been proposed in the present context is as follows. First decompose the array y into 'row effects', 'column effects' and residuals as follows:

$$y_{ij} = \bar{y}_{..} + r_i + c_j + e_{ij},$$

where the row and column sums of e_{ij} are all zero. Then generate bootstrap observations by sampling from the rows, sampling from the columns, and sampling independently from the residuals. That is to say $r^* = r \circ \varphi_r$, $c^* = c \circ \varphi_c$, and $e^* = e \circ \varphi_e$, with $\varphi_r \in \mathscr{M}_r$, $\varphi_c \in \mathscr{M}_c$ and $\varphi_e \in \mathscr{M}_{rc}$. The components in the reconstituted array are

$$y_{ij}^* = \bar{y}_{..} + r_i^* + c_j^* + e_{ij}^*.$$

Note that r^*, c^* and e^* do not satisfy the zero-sum constraint, so there is variation in the mean of the bootstrap samples.

This hybrid scheme is not a bootstrap resampling scheme in the sense that the term is used in this paper because y^* is not obtained from y by monoid composition. Nevertheless, given φ_r, φ_c and φ_e, the transformation from y to y^* is linear. Further, the distributions are uniform on $\mathscr{M}_r$, $\mathscr{M}_c$ and $\mathscr{M}_{rc}$. Consequently, the variance of $\bar{Y}^*$ is a positive symmetric quadratic form in y. By the argument given in the preceding section, the bootstrap variance of $\bar{Y}$ cannot be a consistent estimate of the sampling variance. In fact, the bootstrap

variance given by this hybrid scheme is identical to the Boot-II variance using the product monoid.

4.6. Assumptions

Most of the resampling schemes considered in this paper have the property that the distribution of φ is uniform on some monoid. Uniformity, however, is not critically important: generally speaking the distributions are not uniform on the full monoid. What is important, however, is that the distribution of φ on $\mathcal{M}$ should be independent of y. It is this property that guarantees that the bootstrap moments and cumulants of order r are homogeneous polynomials in y of degree r. If this assumption is violated, the bootstrap variance need not be a quadratic form in y. So it is not impossible that a modified resampling scheme might be devised such that the bootstrap variance is consistent.

To give an example of a resampling scheme in which the critical assumption is violated, suppose we first calculate S_r^2, S_c^2 and S_e^2, the row, column and error sums of squares. Given these values, let Z_i be a sequence of independent Bernoulli variables with parameter $S_e^2/(S_e^2 + S_r^2 + S_c^2)$. If $Z_i = 0$ the ith bootstrap sample is drawn uniformly from the product monoid, otherwise uniformly from the full monoid. The bootstrap distribution is a mixture of Boot-I and Boot-II. The variance in this case is a ratio of homogeneous polynomials of degree 4 and 2. It seems most unlikely that a randomized bootstrap scheme of this sort could yield a consistent variance estimate, but I have been unable to prove this.

5. Pivotal statistics

The two examples discussed above are intended to illustrate a common type of problem in which the target statistic has a natural estimator, $\hat{\mu}$, that is asymptotically normally distributed, but no consistent variance estimate is readily available. The purpose of resampling is then to obtain a variance estimate in order to set approximate confidence intervals for μ. If this can be done successfully, the often difficult analytic exercise of computing a consistent variance estimate is avoided.

When a consistent estimate of the variance, s^2, is available, the nature of the problem is drastically altered. First, the need for bootstrapping is greatly diminished because a first-order approximate confidence interval can be obtained directly from the normal approximation. Second, if resampling is to be used, it is more natural to use the bootstrap distribution of the approximately pivotal statistic $(\bar{Y}^* - \bar{y})/s^*$. The purpose of the bootstrap is then to improve on the first-order normal approximation.

Each family of sampling distributions considered in this paper is characterized by two properties, group invariance and independence of certain components. In the one-way layout, for example, observations in distinct blocks are independent. Monoid composition automatically preserves group invariance, but it need not preserve the independence structure. In the single-sample problem, the permutation distribution is exchangeable but it does not have independent components. In the one-way layout, both group invariance and

independence of blocks are preserved by resampling plans II, V and VI only. For the two-way exchangeable array, both resampling plans have the property that if I, I' are disjoint sets of rows, and J, J' disjoint columns, then Y_{IJ}^* and $Y_{I'J'}^*$ are independent. The hybrid scheme in Section 4.5 has the same property. In each of these resampling schemes, the bootstrap distribution belongs to the family of sampling distributions under consideration. Any identity that is satisfied by the family of sampling distributions is automatically satisfied by the bootstrap distribution.

For the one-way layout, the family of sampling distributions has the property that $nb\,\mathrm{var}(\bar{Y}) = \mathrm{E}(MS_\mathrm{b})$. It follows immediately as a special case for bootstraps II, V and VI, that $nb\,\mathrm{var}(\bar{Y}^*|y) = \mathrm{E}(MS_\mathrm{b}^*|y)$. By the same argument, in the row–column exchangeable array,

$$mn\,\mathrm{var}(\bar{Y}^*|y) = \mathrm{E}(MS_\mathrm{r}^* + MS_\mathrm{c}^* - MS_\mathrm{e}^*|y)$$

for both monoid resampling schemes and the hybrid. As a consequence, for each of these resampling schemes, the studentized statistic $(\bar{Y}^* - \bar{y})/s^*$ is standard normal to first order. Here s^2 is the consistent variance estimate or any asymptotically equivalent statistic. For the pivotal statistic, the bootstrap distributions are in first-order agreement with the sampling distribution of $(\bar{Y} - \mu)/s$. It is unclear if any bootstrap distribution is correct to the next order.

6. Linear regression

6.1. Simple linear regression

Let Ω be the set of units, and let the response vector Y be a random variable in $\mathscr{V} = \mathscr{R}^\Omega$. Given a covariate vector $z \in \mathscr{R}^\Omega$, the expected value of Y satisfies the linear model

$$\mu = \mathrm{E}(Y|z) = \alpha + \gamma z.$$

We draw a distinction between the theoretical residuals $\epsilon = Y - \mu$, which are assumed to be i.i.d., and the fitted residuals $\hat{\epsilon} = Y - PY$, which are not exchangeable. Resampling operates on the theoretical residuals.

For any γ, we define $\varepsilon_\gamma = Y - \gamma z$: if γ is the true value, this vector is exchangeable. Consider the statistic $T = \langle \varepsilon_\gamma, z \rangle_{\mathscr{V}/1}$, the adjusted sum of products $\sum(z_i - \bar{z})\varepsilon_{\gamma i}$, which is also equal to $(\hat{\gamma} - \gamma)\|z\|^2$, where $\|z\|^2 = \sum(z_i - \bar{z})^2$. For resampling purposes, we define $\varepsilon_\gamma^* = \varepsilon_\gamma \circ \varphi$, with φ chosen uniformly from the group or from the monoid. The resampling distribution of T is the distribution of $T^* = \langle \varepsilon_\gamma^*, x \rangle_{\mathscr{V}/1}$, which has mean zero and permutation variance

$$\mathrm{var}(T^*) = \|\varepsilon_\gamma\|^2 \|z\|^2/(n-1) = \|z\|^2((n-2)s^2 + (\hat{\gamma} - \gamma)^2\|z\|^2)/(n-1),$$

where s^2 is the residual mean square after regression on z. If monoid resampling is used, the variance is reduced by the factor $(n-1)/n$.

The preceding argument based on exchangeability differs from standard practice as described by Efron and Tibshirani (1993) or Davison and Hinkley (1997). Although

$T = \langle \varepsilon_\gamma, z \rangle$, these authors consider the reference distribution $\langle \varepsilon_{\hat{\gamma}}^*, z \rangle$, treating the *fitted* residuals as exchangeable. As a partial fix for practical work, Davison and Hinkley recommend resampling from the modified fitted residuals, but these are not exchangeable either.

It is often a reasonable approximation to suppose that the studentized statistic has a fixed distribution independent of γ and z, but dependent on n. In this case, a confidence interval for γ can be obtained as the set

$$\left\{ \gamma : \frac{(\hat{\gamma} - \gamma) \|z\| \sqrt{n-1}}{\sqrt{(n-2)s^2 + (\hat{\gamma} - \gamma)^2 \|z\|^2}} \leqslant C_\alpha \right\}.$$

The usual t ratio is $t = (\hat{\gamma} - \gamma) \|z\| / s$, so the preceding interval coincides with the usual normal-theory interval if

$$C_\alpha = \sqrt{n-1} \, t_{n-2,\alpha} \Big/ \sqrt{n - 2 + t_{n-2,\alpha}^2}.$$

The exact version of the preceding argument requires C_α to be chosen from the empirical distribution of the standardized values

$$\frac{\langle \varepsilon_\gamma^*, z \rangle \sqrt{(n-1)}}{\|\varepsilon_\gamma\| \, \|z\|},$$

with φ chosen uniformly from the group. This exercise must be repeated for a range of γ values in order to construct the required interval. In other words, exact confidence sets are based on the permutation t distribution.

6.2. Multiple regression

In the standard multiple linear regression model, $Y = X\beta + \varepsilon$, the components of the observation vector Y are not exchangeable, but the (theoretical) residual vector ε in $\mathscr{R}^\Omega$ is taken to be i.i.d. and thus exchangeable. Exchangeability here refers to the group of permutations acting on the set of units Ω. Let $\mathscr{X}$ be that p-dimensional subspace of $\mathscr{R}^\Omega$ spanned by the columns of X. What is required is a notion of exchangeability modulo translations in $\mathscr{X}$. One possible definition is as follows. We say that the random variable Y is *exchangeable modulo* $\mathscr{X}$ if there exists a non-random point $x \in \mathscr{X}$ such that the random variable $Y - x$ is exchangeable in $\mathscr{V} = \mathscr{R}^\Omega$. In the standard linear regression model, the observation vector is exchangeable modulo $\mathscr{X}$ because, with $x = X\beta$, $Y - x = \epsilon$ has i.i.d. components. As we will see, this definition is not especially useful for resampling purposes because the point x is typically unknown.

Consider now an extension of the linear regression model

$$\mathrm{E}(Y) = X\beta + z\gamma$$

in which z is a given vector in $\mathscr{R}^\Omega$, but not in $\mathscr{X}$. A confidence interval is required for γ under the assumption that $Y - z\gamma$ is exchangeable modulo $\mathscr{X}$. For purposes of testing and estimation, we choose the conventional statistic, the quotient-space inner product,

$$T = \langle Y, z \rangle_{\mathscr{V}/\mathscr{X}} = \langle z\gamma + \epsilon, z \rangle_{\mathscr{V}/\mathscr{X}} = \hat{\gamma} \langle z, z \rangle_{\mathscr{V}/\mathscr{X}},$$

in which all vectors are regarded as elements in the quotient space $\mathscr{R}^{\Omega}$ modulo $\mathscr{X}$. In the more familiar matrix notation

$$T = Y^{\mathrm{T}} W (I - P) z,$$

where $P = X(X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W$ is the orthogonal projection on to $\mathscr{X}$. In the present context of exchangeable errors, the inner product matrix must be invariant. We assume that the off-diagonal elements are zero, so $W = I$.

For resampling purposes, it is inappropriate to use $Y - z\gamma$ directly because its components are far from exchangeable. Nor can we use the the quotient space version $Y - z\gamma + \mathscr{X}$ because this is not a point in $\mathscr{R}^{\Omega}$. The only remaining option is to use the projection $\hat{\varepsilon}_{\gamma} = (I - P)(y - z\gamma)$ onto $\mathscr{X}^{\perp}$. However, the residual vector thus defined does not satisfy the definition of exchangeability modulo $\mathscr{X}$. Nevertheless, we take $\hat{\epsilon}_{\gamma}$ as given, and proceed as if these residuals are exchangeable. That is to say, we construct resampled residuals $\varepsilon_{\gamma}^{*} = \hat{\varepsilon}_{\gamma} \circ \varphi$, where φ is chosen uniformly at random either from the symmetric group or the monoid on Ω. Then, for each γ, we compare the observed observed value of T_{γ}

$$t_{\mathrm{obs}}(\gamma) = \hat{\varepsilon}_{\gamma}^{\mathrm{T}} (I - P) z = (\hat{\gamma} - \gamma) z^{\mathrm{T}} (I - P) z$$

with the bootstrap distribution of values

$$T_{\gamma}^{*}(\varphi) = (\hat{\varepsilon}_{\gamma} \circ \varphi)^{\mathrm{T}} (I - P) z$$

for φ in $\mathscr{G}$ or in $\mathscr{M}$. The set of values of γ for which $t_{\mathrm{obs}}(\gamma)$ does not fall in the tails of the relevant distribution constitutes a confidence set for γ, although the coverage probability may not coincide with the nominal level.

On the assumption that the constant functions are in $\mathscr{X}$, that is, that the model X includes the intercept, and that the resampling distribution is uniform on the group, the bootstrap mean of T^{*} is zero and the variance is

$$\mathrm{var}(T_{\gamma}^{*} | \hat{\varepsilon}_{\gamma}) = \|\hat{\varepsilon}_{\gamma}\|^{2} \|z\|^{2} / (n - 1)$$

$$= ((n - p - 1)s^{2} + (\hat{\gamma} - \gamma)^{2} \|z\|^{2}) \|z\|^{2} / (n - 1),$$

where $\|z\|^{2} = z^{\mathrm{T}} (I - P) z$, and s^{2} is the residual mean square after regression on both X and z. In the more familiar normal-theory framework, T is normally distributed with mean zero and variance $\sigma^{2} \|z\|^{2}$ under the hypothesis. By comparison, the resampling variance is too small by the factor $(n - p)/(n - 1)$.

If we make the approximation that the resampling distribution is normal, the bootstrap interval for γ becomes

$$\left\{ \gamma : \frac{(n - 1)(\hat{\gamma} - \gamma)^{2} \|z\|^{2}}{(n - p - 1)s^{2} + (\hat{\gamma} - \gamma)^{2} \|z\|^{2}} \leq \chi_{1,\alpha}^{2} \right\}.$$

In terms of the more familiar F ratio, $F = (\hat{\gamma} - \gamma)^{2} \|z\|^{2} / s^{2}$, on $(1, n - p - 1)$ degrees of freedom, the bootstrap interval is

$$\left\{ \gamma : \frac{(n-1)F}{n-p-1+F} \leqslant \chi^2_{1,a} \right\}.$$

If sampling is done uniformly on the monoid, i.e. with replacement, the set is

$$\left\{ \gamma : \frac{nF}{n-p-1+F} \leqslant \chi^2_{1,a} \right\}.$$

Both intervals are too narrow, and could be improved by multiplicative adjustment of the bootstrap variance or by adjustment of the residuals as recommended by Davison and Hinkley (1997). But even with this correction, the coverage probability does not achieve the nominal level.

6.3. Rotational exchangeability

The preceding derivation is unsatisfactory because it was assumed that $(I-P)(Y-\gamma z)$ is exchangeable in $\mathscr{R}^\Omega$. This assumption is clearly false because the residuals are known to lie in the subspace $\mathscr{X}^{\cdot\perp}$. Unless $\mathscr{X}^{\cdot}$ is an invariant subspace, that is, either $\mathscr{X}^{\cdot} = 0$, 1, or $1^\perp$, the residuals cannot be exchangeable. The difficulty arises from the fact that the resampled residuals have a component in $\mathscr{X}^{\cdot}$, which does not contribute to the statistic T^*.

One possible remedy is to replace $\varepsilon^* = \hat{\varepsilon} \circ \varphi$ by

$$\varepsilon' = Q\varepsilon^* \|\hat{\varepsilon}\| / \|Q\varepsilon^*\|,$$

where $Q = I - P$. In other words, the resampled residuals are projected onto $\mathscr{X}^{\cdot\perp}$ and rescaled to have the same norm as $\hat{\varepsilon}$. Although similar in spirit, this remedy differs from Davison and Hinkley's (1997, p. 259) recommendation of using modified residuals for the same purpose. A serious drawback with e' is that if there exists a φ such that $\varepsilon^* \in \mathscr{X}^{\cdot}$, $Q\varepsilon^*$ is zero and ε' is undefined. The existence of such a φ is assured if monoid resampling is used. If resampling is restricted to the group, however, there does not ordinarily exist a $\varphi \in \mathscr{G}$ such that $Q\varepsilon^* = 0$. One exception, however, occurs if $\mathscr{X}^{\cdot}$ is a factorial model for a 2^k factorial design in which $p \geqslant n - k$.

An alternative, but closely related, option is to consider the subgroup of $GL(\mathscr{V}, \mathscr{V})$ that leaves the subspace $\mathscr{X}^{\cdot}$ invariant. It is enough in fact to consider the group $\mathscr{G}_\mathscr{X}$ of orthogonal transformations on $\mathscr{V}$ that leaves $\mathscr{X}^{\cdot}$ invariant. We say that Y is spherically symmetric modulo $\mathscr{X}^{\cdot}$ if, for each $\varphi \in \mathscr{G}_\mathscr{X}$, the distribution of φY is the same as the distribution of Y. Invariance under $\mathscr{G}_\mathscr{X}$ is a weaker assumption than spherical symmetry for the distribution of ε in $\mathscr{V}$. For example, in a blocked design, spherical symmetry modulo block effects is usually more sensible than full spherical symmetry. Apart from the normal-theory model, full spherical symmetry is rarely an appealing assumption.

We proceed on the basis of $\mathscr{G}_\mathscr{X}$-exchangeability modulo $\mathscr{X}^{\cdot}$. Since $\mathscr{X}^{\cdot}$ is invariant, we operate directly on the vector $\hat{\varepsilon}_\gamma = Y - \gamma_0 z$ by an element φ in $\mathscr{G}_\mathscr{X}$. By assumption, the rotated vector $\varepsilon^* = \varphi \hat{\varepsilon}_\gamma$ has the same distribution as $\hat{\varepsilon}_\gamma$. The calculations proceed as above except that φ is chosen uniformly with respect to Harr measure on $\mathscr{G}_\mathscr{X}$. This amounts to saying that ε^* is uniformly distributed on the surface of the sphere of radius $(\hat{\varepsilon}_\gamma^T (I-P) \hat{\varepsilon}_\gamma)^{1/2}$ in $\mathscr{V}/\mathscr{X}^{\cdot}$. In this case, the bootstrap distributional calculations coincide

with conditional normal-theory calculations for the conditional distribution of T given the sufficient statistics $X^T Y$ and s_0^2. So the one-sided bootstrap confidence interval for γ is

$$(-\infty, \hat{\gamma} + t_{n-p-1,a} s / \|z\|),$$

where $\|z\|$ is the quotient space norm $(z^T Q z)^{1/2}$.

The preceding argument is closely related to Box (1976, Section 3.10) and Dawid (1977) who note that rotational symmetry is sufficient to derive the traditional t and F statistics for linear models.

Note that, since $\mathscr{G}_{\mathscr{X}}$. is infinite, the distinction between sampling with and without replacement does not arise.

7. Conclusions

The main purpose of this paper is to test the limits of resampling as an inferential technique. Under what conditions is the bootstrap distribution of a statistic a faithful estimate of the sampling distribution? The question has been investigated by studying the distributions of linear statistics computed on exchangeable arrays. It is found that, for non-pivotal statistics, the bootstrap distribution is ordinarily not a consistent estimate of the sampling distribution. For the two-way exchangeable array, no resampling scheme exists such that the bootstrap variance of $\bar{Y}^*$ consistently estimates the sampling variance of this statistic. Outside of the i.i.d. case, it appears that resampling is not a reliable method for obtaining a consistent variance estimate.

Acknowledgement

References

Box, G.E.P. (1976) Science and statistics. *J. Amer. Statist. Assoc.*, **71**, 791–799.

Davison, A. and Hinkley, D.V. (1997) *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Dawid, A.P. (1977) Spherical matrix distributions and a multivariate model. *J. Roy. Statist. Soc. Ser. B*, **39**, 254–261.

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. London: Chapman & Hall.

Fisher, R.A. (1935) *Design of Experiments*. Edinburgh: Oliver and Boyd.

McCullagh, P. (1987) *Tensor Methods in Statistics*. London: Chapman & Hall.