Kernels of MVU Moment Estimators in Balanced Designs

D. S. Robson

BU-373-M

June, 1971

Abstract

The order statistic $\left\{X_1,\ldots,X_n\right\}$ of a simple random sample $\left(X_1,\ldots,X_n\right)$ is complete and minimal sufficient with respect to the class of all possible distribution functions $F(x)$. The unique MVUE of any polynomial function of the moments of F may therefore be generated by constructing a kernel estimator (an unbiased estimator which is a function of as few of the n observations as possible) and calculating the conditional expected value of this kernel, given the order statistic $\left\{X_1,\ldots,X_n\right\}$. This well known method of constructing MVUE moment estimators may be applied (in a less well known manner) to the more complicated problem of estimating moments such as variance components from an array of observations generated by a balanced experimental design.

Kernels of MVU Moment Estimators in Balanced Designs

D. S. Robson

BU-373-M June, 1971

Introduction

It is well known that for Model II balanced designs the MVUE variance component estimators derived for the conventional, homoscedastic normal model are also MVUE with respect to the general Model II in which distributions are unspecified. Thus, if a balanced sample is drawn from a distribution $f_{XYZ}(x,y,z) = f_X(x)f_Y(y)f_{Z|XY}(z|x,y)$ by independently selecting r observations from $f_X(x)$, c from $f_Y(y)$ and n from each of the rc distributions $f_{Z|X,Y}(z|x,y)$ determined by the r × c factorial array of (x,y) observations, then conventional unbiased estimators of the variance components of Z are MVUE with respect to the class $\mathfrak{F}$ of all possible distributions $f_{XYZ}(x,y,z)$ in which X and Y are statistically independent. The error mean square

$$s^2 = \frac{1}{rc}\sum_i \sum_j \frac{\sum_{k=1}^{n}(Z_{ijk} - \bar{Z}_{ij.})^2}{n-1}$$

for example is the MVUE of

$$\sigma_\epsilon^2 = E\left[Z - E(Z|X,Y)\right]^2$$

The parameter σ_ϵ^2 is of <u>degree</u>* 2 since a minimum of 2 observations is required to estimate σ_ϵ^2. A <u>kernel</u> of s^2 is any function of 2 of the nrc observations which is an unbiased estimator of σ_ϵ^2. The array of nrc observations on Z may be arranged in

* The concept of degree of a parameter is ill formulated here and represents only total degree.

a two-way table of r rows and c columns with n observations per cell, and the order statistic of this array is the collection of $(n!)^{rc} r!c!$ arrays generated by permuting rows, permuting columns, and permuting the observations in each cell. If the random variables $(X_1,\ldots,X_r)$, $(Y_1,\ldots,Y_c)$ are unobservable* then this order statistic of the Z-array is complete and minimal sufficient with respect to the class $\mathfrak{F}$, and the (unique) MVUE of any moment parameter such as σ_ϵ^2 may then be obtained by constructing a kernel estimator and averaging this over all $(n!)^{rc} r!c!$ permutations of the Z-array. The kernel $Z_{111}^2 - Z_{111}Z_{112}$, for example, is an unbiased estimator of σ_ϵ^2; averaging this kernel over all permutations of the n observations $Z_{111},\ldots,Z_{11n}$ in cell (1,1) gives

$$\frac{1}{n}\sum_{k=1}^{n} Z_{11k}^2 - \frac{1}{n(n-1)}\sum_{k\neq k'} Z_{11k}Z_{11k'} = \frac{1}{n-1}\sum_{k=1}^{n}(Z_{11k} - \bar{Z}_{11\cdot})^2 \equiv s_{ij}^2$$

and averaging s_{ij}^2 over all permutations of rows and columns gives

$$s^2 = \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c} s_{ij}^2 \quad .$$

We present here some other examples of the use of kernels in constructing MVU moment estimators.

* If the statistical problem concerns estimation of moments of Z only then the question of whether X and Y are observable or unobservable is irrelevant in this general case where distributions are unspecified.

Examples

We shall restrict examples to the balanced two-way array described above since extensions to higher-way arrays will be obvious. To pose the problem in more familiar terms let

$$\mu \equiv E(Z)$$

$$\rho_i \equiv E(Z|X_i) - \mu$$

$$\tau_j \equiv E(Z|Y_j) - \mu$$

$$\gamma_{ij} \equiv E(Z|X_i,Y_j) - (\mu + \rho_i + \tau_j)$$

$$\epsilon_{ijk} \equiv Z_{ijk} - (\mu + \rho_i + \tau_j + \gamma_{ij})$$

and to begin with a familiar example consider the problem of estimating the variance component σ^2_γ ,

$$\sigma^2_\gamma \equiv E\left[E(Z|X_i,Y_j) - E(Z|X_i) - E(Z|Y_j) + E(Z)\right]^2$$

$$\equiv E\gamma^2_{ij}$$

Noting that

$$Z_{111} - Z_{121} - Z_{211} + Z_{221} \equiv (\gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22}) + (\epsilon_{111} - \epsilon_{121} - \epsilon_{211} + \epsilon_{221})$$

is correlated with

$$Z_{112} \equiv \mu + \rho_i + \tau_1 + \gamma_{11} + \epsilon_{112}$$

only through the common element γ_{11} we see that

$$EZ_{112}\left(Z_{111} - Z_{121} - Z_{211} + Z_{221}\right) = E\gamma_{11}^2 = \sigma_\gamma^2$$

and thus we have found a kernel (of degree 5) estimator of σ_γ^2. Averaging this kernel over within-cell permutations gives

$$\frac{1}{n(n-1)} \sum_{k \neq k'} Z_{11k} Z_{11k'} - \bar{Z}_{11\cdot}\bar{Z}_{12\cdot} - \bar{Z}_{11\cdot}\bar{Z}_{21\cdot} - \bar{Z}_{11\cdot}\bar{Z}_{22\cdot}$$

$$= \frac{1}{n(n-1)} \left[\left(\sum_1^n Z_{11k}\right)^2 - \sum_1^n Z_{11k}^2 \right] - \bar{Z}_{11\cdot}\bar{Z}_{12\cdot} - \bar{Z}_{11\cdot}\bar{Z}_{21\cdot} - \bar{Z}_{11\cdot}\bar{Z}_{22\cdot}$$

Averaging over permutations of rows and columns gives

$$\frac{1}{nrc(n-1)} \sum_{i=1}^r \sum_{j=1}^c \left[\left(\sum_{k=1}^n Z_{ijk}\right)^2 - \sum_{k=1}^n Z_{ijk}^2 \right] - \frac{1}{c-1} \left[\frac{c}{r} \sum_{i=1}^n \bar{Z}_{i\cdot\cdot}^2 - \frac{1}{rc} \sum_1^r \sum_1^c \bar{Z}_{ij\cdot}^2 \right]$$

$$- \frac{1}{r-1} \left[\frac{r}{c} \sum_1^c \bar{Z}_{\cdot j\cdot}^2 - \frac{1}{rc} \sum_1^r \sum_1^c \bar{Z}_{ij\cdot}^2 \right]$$

$$+ \frac{1}{(r-1)(c-1)} \left[rc\bar{Z}_{\cdot\cdot\cdot}^2 - \frac{c}{r} \sum_1^r \bar{Z}_{i\cdot\cdot}^2 - \frac{r}{c} \sum_1^c \bar{Z}_{\cdot j\cdot}^2 + \frac{1}{rc} \sum_1^r \sum_1^c \bar{Z}_{ij\cdot}^2 \right]$$

which simplifies to the familiar form

$$\hat{\sigma}^2_\gamma = \frac{1}{(r-1)(c-1)}\left[\sum_1^n \sum_1^c Z^2_{ij\cdot} - c\sum_1^r \bar{Z}^2_{i\cdot\cdot} - r\sum_1^c \bar{Z}^2_{\cdot j\cdot} + rc\bar{Z}^2_{\cdot\cdot\cdot}\right]$$

$$- \frac{1}{n(n-1)}\left[\sum_1^r \sum_1^c \sum_1^n Z^2_{ijk} - n\sum_1^r \sum_1^c \bar{Z}^2_{ij\cdot}\right]$$

As a second example we shall construct a MVUE of the covariance between cell means and within-cell variances -- i.e., an estimator of

$$\mathrm{Cov}\left[E(Z|X,Y),\ E(Z^2|X,Y) - \{E(Z|X,Y)\}^2\right]$$

$$= \mathrm{Cov}\left(\mu + \rho_i + \tau_j + \gamma_{ij},\ \sigma^2_{ij}\right) \equiv \mathrm{Cov}\left(\mu_{ij},\ \sigma^2_{ij}\right)$$

where

$$\sigma^2_{ij} = E(Z^2|X_i,Y_j) - \{E(Z|X_i,Y_j)\}^2$$

A kernel for σ^2_{ij} is $Z^2_{ijk} - Z_{ijk}Z_{ijk'}$, and a kernel for the product $\mu_{ij}\sigma^2_{ij}$ is then $Z_{ijk''}\left(Z^2_{ijk} - Z_{ijk}Z_{ijk'}\right)$ while a kernel for $\mu\sigma^2_\epsilon$ is $Z_{i'j'k}\left(Z^2_{ijk} - Z_{ijk}Z_{ijk'}\right)$. A kernel (of total degree 4) for the covariance is then

$$Z^2_{ijk}Z_{ijk'} - Z_{ijk}Z_{ijk'}Z_{ijk''} - Z^2_{ijk}Z_{i'j'k} + Z_{ijk}Z_{ijk'}Z_{i'j'k} .$$

Averaging over within-cell permutations gives

$$\frac{1}{n(n-1)}\left[\sum_k Z_{ijk} \sum_k Z^2_{ijk} - \sum_k Z^3_{ijk}\right] - \frac{1}{n(n-1)(n-2)}\left[\left(\sum_k Z_{ijk}\right)^3 - 3\sum_k Z_{ijk}\sum_k Z^2_{ijk} + 2\sum_k Z^3_{ijk}\right]$$

$$-\frac{1}{n^2}\sum_k Z^2_{ijk}\sum_k Z_{i'j'k} + \frac{1}{n^2(n-1)}\left[\left(\sum_k Z_{ijk}\right)^2 - \sum_k Z^2_{ijk}\right]\sum_k Z_{i'j'k}$$

Averaging over permutations of rows and columns gives

$$\frac{1}{nrc(n-1)(n-2)}\left[(n+1)\sum_i\sum_j\left(\sum_k Z_{ijk}\right)\left(\sum_k Z^2_{ijk}\right) - n\sum_i\sum_j\sum_k Z^3_{ijk} - \sum_i\sum_j\left(\sum_k Z_{ijk}\right)^3\right]$$

$$-\frac{1}{(n-1)(r-1)(c-1)}\left[rc\left(\sum_i\sum_j\sum_k Z_{ijk}\right)\sum_i\sum_j\sum_k\left(Z_{ijk} - \bar{Z}_{ij.}\right)^2\right.$$

$$-\frac{c}{r}\sum_i\left(\sum_j\sum_k Z_{ijk}\right)\sum_j\sum_k\left(Z_{ijk} - \bar{Z}_{ij.}\right)^2 - \frac{r}{c}\sum_j\left(\sum_i\sum_k Z_{ijk}\right)\sum_i\sum_k\left(Z_{ijk} - \bar{Z}_{ij}\right)^2$$

$$\left.+\frac{1}{rc}\sum_i\sum_j\left(\sum_k Z_{ijk}\right)\sum\left(Z_{ijk} - \bar{Z}_{ij.}\right)^2\right]$$

The algebra of this approach to moment estimation clearly becomes very tedious for moments of higher degree, however, there is little interest in moments beyond variances of variances. An algebra of symmetric functions is available and does become useful beyond this limit.