

## On Use of Bootstrapping for Estimation of Confidence Interval of Heritability

V.K. Bhatia and J. Jayasankar\*

*IASRI, New Delhi*

(Received : September, 1994)

### SUMMARY

This work is aimed at the confidence interval estimation of heritability using the non-parametric bootstrap procedure. Besides estimation of confidence intervals, estimates of bias and subsequently bias corrected confidence intervals have also been obtained. The superiority and the ease of bootstrap technique is highlighted with the help of simulated and real data.

*Key words* : Heritability, Intra-class correlation, Confidence interval, Bootstrap, Bias correction.

### 1. Introduction

Heritability is one of the most used and highly dependable genetic parameter which measures the quantitative inheritance of economic trait. This particular population parameter has been exploited to the hilt in the construction of selection indices. So naturally the precision and accuracy with which heritability is estimated have a big say in such ventures. As far as estimation is concerned a set of standard procedures is available, either as a function of intra-class correlation or as intra-group regression (Falconer [8]). But when it comes to the estimation of precision no exact expression is available but for those based on Fisher's expression for intra-class correlation or by some linearization methods like that of Taylor's (Falconer, [8]).

The chief drawback of the above procedures is that they are highly approximated and are based on such assumptions as normality of the observations. When it comes to the confidence interval estimation no serious efforts seem to have been taken before 1980's. In the early 80's efforts were made to consider heritability as a function of ratio of two mutually orthogonal variance components and under the normality assumptions F - distribution was used (Tai, [13]). In another work this has been viewed through a mixed linear model problem (Harville and Fenech [10]). With these limitations a

\* Present address : CSWRI, Avikanagar, Malpura (Raj)

non-parametric, heuristic, less theoretical and analytical procedure, has been thought of as an ideal one, and the resampling procedure, called, the Bootstrap has been exploited. Though this bootstrap seems to be more like a technique developed out of necessity as a stop-gap arrangement, it has sound theoretical base too (Efron [3], [4], [5]; Efron and Tibshirani [6], [7]; Hall [9]). The most striking feature of this resampling technique is generality towards diversified problems. Moreover it can be carried out for regression models too (Aast Veit [1]; Bhatia *et al.* [2]).

In this investigation an attempt has been made to estimate the confidence interval along with variance and bias of heritability estimators using the non-parametric bootstrap technique.

## 2. Methodology

The estimate of heritability in case of half-sib analysis is simply four times the intra-class correlation (Falconer [8]).

i.e.

$$\hat{h}^2 = 4 \times \hat{t}$$

where  $\hat{t}$  is the intra-class correlation and  $\hat{h}^2$  is estimate of heritability. Then the approximate variance estimate of the heritability estimator is given as

$\hat{\sigma}_{\hat{h}^2}^2 = 16 \hat{\sigma}_{\hat{t}}^2$  where  $\hat{\sigma}_{\hat{t}}^2$  is the variance estimate of the intra-class correlation and approximately

$$\hat{\sigma}_{\hat{t}}^2 = \frac{8 \hat{t}}{m n}$$

with  $m$  as number of half-sib families and  $n$  as size of family,

which means that

$$\hat{\sigma}_{\hat{h}^2}^2 = \frac{32 \hat{h}^2}{(m n)} \text{ approximately.}$$

Another way of estimating precision is by the Fisher's Z transformation method of intra-class correlation which has been done for comparison sake in this investigation.

## 3. Bootstrapping

In this work bootstrapping is applied at the sire level and the progeny structure per sire is kept intact. That is from the data of  $m$  sires each having progeny of size  $n$ , it has so arranged that once the  $i$ th sire is selected all the  $n$  progeny attached to it are automatically selected and a new bootstrap sample

is thus got. Sires are selected at random with replacement by a sequence of random numbers. From each of the bootstrap samples bootstrap replication of heritability is calculated and the parameters of interest are estimated by the following formulae :

Let  $h^2$  be the parameter of interest. Let  $\hat{h}^2$  be the estimator based on the original data. Let  $h^{2*} (1), h^{2*} (2) \dots h^{2*} (N)$  be the  $N$  bootstrap replications.

Then  $E(\hat{h}^2)$  is estimated by

$$\hat{h}^{2*} (.) = \frac{1}{N} \sum_{i=1}^N \hat{h}^{2*} (i)$$

The bias is estimated by

$$\hat{B}_{\hat{h}^2} = \hat{h}^{2*} (.) - \hat{h}^2$$

The sample standard deviation is estimated by

$$\hat{SE}_{\text{BOOT}}^* = \left[ \frac{1}{N-1} \sum_{i=1}^N \left( \hat{h}^{2*} (i) - \hat{h}^{2*} (.) \right)^2 \right]^{1/2}$$

The approximate confidence intervals or more popularly percentile intervals are computed by

$$h^2 \in [\hat{G}^{-1} (\alpha/2), \hat{G}^{-1} (1 - \alpha/2)]$$

where  $1 - \alpha$  is the level of confidence and  $G (s) = \hat{P}_* (\hat{h}^{2*} < s)$  where  $\hat{P}_*$  indicates the probability computed according to the bootstrap distribution of  $\hat{h}^{2*}$  (Efron [4]). But heritability being a non-linear parameter (ratio of two variances), is expected to be estimated with a significant amount of bias. So it becomes mandatory to correct with some bias correction measure. Following the line of Efron [4], the bias-corrected percentiles are calculated in the following way

Let  $Z_o = \Phi^{-1} (\hat{G} (\hat{h}^2))$  where  $\Phi^{-1}$  is the inverse function of the standard normal c.d.f. Then the bias corrected percentiles method ends up with an approximate  $(1 - \alpha)$  confidence interval for  $h^2$  given by

$$\left\{ \hat{G}^{-1} [\Phi (2z_o - z_{\alpha/2})], \hat{G}^{-1} [\Phi (2z_o + z_{\alpha/2})] \right\}$$

For the intervals the following three characteristics are worked out for comparison sake (Efron [5]).

- (i) Length = upper limit of the confidence interval - lower limit of the confidence interval.
- (ii) Normalised length = length /  $(2 \times z_{\alpha/2} \times \hat{SE}_{\text{BOOT}})$
- (iii) Shape =  $\log_e \left| \frac{\hat{G}(1 - \alpha/2) - \hat{G}(0.5)}{\hat{G}(0.5) - \hat{G}(\alpha/2)} \right|$

#### 4. Results and Discussion

In order to study the performance of variance, bias and confidence interval over repeated bootstrapping, two types of data are used. The first being the simulated one and the second a live-data set.

##### (a) Simulated Data

The simulation is carried out by constructing three populations as per half-sib model (Ronningen [12]) with parameters of heritability low, intermediate and high as 0.1, 0.5 and 0.8 respectively with 20 sires and each having progeny size 7. Bootstrap samples were got with the help of a random number generator. For each sample bootstrapping was done with 100, 500, 1000 and 2000 resamples and parameters of interest were worked out. The results are given in Table 1.

Table 1.

No. of samples	Mean	Median	SE	Bias	Percentile Interval (95%)
Population parameter 0.1					
		$\hat{h}^2 = 0.1164$	$\hat{SE}(\hat{h}^2) = 0.2284$		
100	0.0599	0.0283	0.2370	0.0565	(-0.4635, 0.5186)
500	0.0789	0.0769	0.2329	0.0375	(-0.4077, 0.5610)
1000	0.0814	0.0845	0.2366	0.0350	(-0.4015, 0.5524)
2000	0.0754	0.0751	0.2377	0.0410	(-0.3999, 0.5508)
Population parameter 0.5					
		$\hat{h}^2 = 0.4405$	$\hat{SE}(\hat{h}^2) = 0.2959$		
100	0.4382	0.4284	0.3918	0.0023	(-0.2963, 1.1512)
500	0.3961	0.4145	0.3930	0.0444	(-0.3922, 1.1295)
1000	0.3974	0.4099	0.3902	0.0431	(-0.4144, 1.1484)
2000	0.3625	0.3756	0.3835	0.0780	(-0.4144, 1.1484)
Population parameter 0.8					
		$\hat{h}^2 = 0.8706$	$\hat{SE}(\hat{h}^2) = 0.3613$		
100	0.8057	0.8042	0.2799	0.0649	(0.2459, 1.3655)
500	0.8179	0.8256	0.3118	0.0527	(0.1943, 1.4414)
1000	0.8280	0.8337	0.3114	0.0426	(0.2052, 1.4508)
2000	0.8191	0.8283	0.3100	0.0515	(0.1991, 1.4391)

Data generated with 0.1 as population heritability value gave a sample with estimate 0.1164 and standard error 0.2284. From the table it can be noticed that the estimate of the expected value viz. the mean gets stabilized with 100 - 500 resamples to the range of 0.079 and so is the case of standard deviation which converges around 0.23 right from 100 level itself. The beta values were also obtained which clearly indicated the skewed and platykurtic nature of the distribution of heritability estimator. The percentile interval is asymmetrical and is distinctly shorter than the symmetric theoretical one ( $0.1164 + 2 \times 0.2284$ ).

The generated sample for the population value of 0.5 had an estimated value of 0.4405 and a standard error as 0.2959. It can be easily noticed that apart from the features previously discussed, the median happens to be closer to the sample value than the mean and the standard error being under estimated. As in the previous case the curve of heritability estimator's distribution is skewed and platykurtic. The length is more, as far as the percentile intervals are concerned, than that of the normal theory symmetric interval but more realistic in wake of the bias and beta values.

Just like the first case the third is also an extreme case with data being generated for an heritability value of 0.8. The sample estimate is 0.8706 with a standard error of 0.3613. The mean settles in the range of 0.81 and median slightly more than that. The distribution is gently skewed and platykurtic as it had been in the previous cases. Due to the gentle skewness in the present case the theoretical and percentile intervals are very close.

#### (b) Real Data

The data involving 5 sires and 9 offspring per sire as given in Narain *et al.* [11] is taken into consideration. The estimate of heritability obtained from this data is 0.57682 with an estimate of standard error as 0.3257. The bootstrap mean, standard error and median etc. are given in Table 2.

The trend is strikingly similar to the simulated case, with median being closer to the sample estimate than the mean. Another striking feature is the astonishing closeness of bootstrap standard error with the theoretical one. But still we can see from Tabel 2 that 95% confidence intervals are not symmetric. One more feature noteworthy is the non-fixation of the limits at 100, 500 or 1000 resampling levels. With the bias being significant (bias = 0.1822), bias-correction becomes logical and the bias corrected 95% intervals are given in Table 3.

It is observed that the bias-correction had significant impact with a significant reduction in the normalized length and more stability in the shape. It is thus concluded that for reliable confidence intervals, the bias correction is very important. From the estimate of shape, it is observed that for 2000 bootstrap resamples it becomes nearly symmetrical as the value is nearing 0.5.

Table 2. Uncorrected Confidence Intervals (Real Data (Narain et. al. [11]))

	Lower	Upper	Length	Normalised Length	Shape
N = 100 mean = 0.3860 median = 0.4465 bias = 0.1822 s.e. = 0.3695	-0.4746	0.9153	1.3899	0.9594	-0.6754
N = 500 mean = 0.3770 median = 0.3771 bias = 0.1926 s.e. = 0.3377	-0.4290	0.9892	1.4183	1.0713	-0.2752
N = 1000 mean = 0.3705 median = 0.3640 bias = 0.1926 s.e. = 0.3261	-0.4147	0.9892	1.4040	1.0981	-0.2194
N = 2000 mean = 0.3771 median = 0.3770 bias = 0.1910 s.e. = 0.3241	-0.3780	0.9823	1.3602	1.0705	-0.2210

Table 3. Bias-Corrected Confidence Intervals

N	Interval	Length	Normalised Length	Shape
100	[-0.0878, 0.9275]	1.0153	0.8350	0.9001
500	[0.1775, 1.0290]	0.8515	0.6432	1.1835
1000	[0.0912, 1.0237]	0.9325	0.7294	0.8834
2000	[-0.0071, 1.0200]	1.0271	0.8083	0.5152

Finally from both the data sets, simulated and real, it is noticed that the bias is present. The distribution of heritability estimator is skewed and platykurtic, contrary to the normality assumptions used in the interval estimation. It is also observed that as far as standard error estimation is concerned less than 500 bootstrap re-samples are sufficient as variance stabilizes by that range, but for confidence interval estimation, more than 1000 resamples are needed.

To sum up, this bootstrap technique, though computer intensive, is handy and pliable for the study of the distributional properties of complicated estimators like heritability estimator and also provides the exact estimates of precision and reliability.

## REFERENCES

- [1] Aastveit, A.H., 1990. Use of bootstrapping for estimation of standard deviation and confidence intervals of genetic variance and variance component. *Biom. J.*, **34**, 515-517.
- [2] Bhatia, V.K., Jayasankar, J. and Wahi, S.D., 1994. Use of bootstrap technique for variance estimation of heritability estimators. *Ann. Agric. Res.*, **15**, 476-480.
- [3] Efron, R., 1979. Bootstrap methods : Another look at Jackknife. *Ann. Statist.*, **7**(1), 1-26.
- [4] Efron, B., 1982. The Jackknife, the Bootstrap and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph, 38, Philadelphia : SS/M.
- [5] Efron, B., 1992. Jackknife-after-Bootstrap standard errors and influence functions. *J.R. Statist. Soc.B*, **54**, 83-127.
- [6] Efron, B. and Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, **1**, 54-77.
- [7] Efron, B. and Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall.
- [8] Falconer, D.S., 1989. An Introduction to Quantitative Genetics. E.L.B.S., London.
- [9] Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer Verlag.
- [10] Harville, D.A. and Fenech, A.P., 1985. Confidence intervals for variance ratio, or for heritability, in an unbalanced mixed linear model. *Biometrics*, **41**, 137-152.
- [11] Narain, P., Bhatia, V.K. and Malhotra, P.K., 1979. Handbook of Statistical Genetics : IASRI Publication.
- [12] Ronningen, K., 1974. Monte-Carlo simulation of statistical biological models which are of interest in animal breeding. *Acta Agricultural Scandinavica*, **24**, 135-142.
- [13] Tai, G.C.C., 1983. Confidence intervals of genetic parameters estimated from progeny test experiments in plant breeding. *Can. J. Genet. Cytol.*, **25**, 502-508.