Statistics and Applications {ISSN 2452-7395 (online)}
Volume 15 Nos. 1&2, 2017 (New Series), pp 167-171

Some Refinements of Block Total Response Technique in the context of RRT Methodology

Bikas K Sinha
Indian Statistical Institute, Kolkata [Retired Professor]

Abstract

In this article we propose to discuss some further aspects of Block Total Response Techniques [BTRTs] in the framework of Randomized Response Technique [RRT]. The purpose is to be able to elicit 'truthful' response on a *sensitive* feature from the sampled respondents (of a finite labeled population of respondents), so that eventually the population proportion of incidence of a sensitive feature [SQlF] can be unbiasedly estimated. Towards this, a novel technique was introduced by Raghavarao and Federer (1979) and it was termed 'Block Total Response' [BTR] Technique. In Nandy et al. (2016), we undertook various meaningful versions/generalizations of the BTR Tech-nique, after a brief review of the literature in this direction. In the process, we also introduced Empirical Bayes [EB] estimators.

Keywords: Sensitive feature, Regular feature, Randomized response technique, Block total response technique, Binary proper equireplicate block design, Supplementary block design.

1 Randomized Response Technique [RRT]

We refer to an excellent expository early book on RRT by Chaudhuri and Mukerjee (1988). Hedayat and Sinha (1991), Chapter 11, also provides a fairly complete account of RRT. Two most recent books [Chaudhuri (2011) and Chaudhuri and Christofides (2013)] are worth mentioning as well.

In this paper, we confine our discussion to one Sensitive Qualitative Feature [SQlF] denoted by Q^*, with a binary response [yes/no] from each respondent in the sample/population and we denote by P^*, the corresponding population proportion(s) of 'yes' response. The problem is to provide (i) a method of ascertaining 'truthful' response to the Q^* from each respondent facing the SQlF in the surveyed population and (ii) to provide unbiased estimator of P^*. Generally, simple random sampling with replacement of respondents is contemplated.

2 Block Total Response Technique [BTRT]

Raghavarao and Federer (1979) introduced BTRT. A precursor to this study was undertaken by Smith et al. (1974). Here is an informative reference: Nayak and Adeshiyan (2009).

For completeness, we describe the basic idea first.

Corresponding Author: Bikas K Sinha
Email id:- bikassinha1946@gmail.com

Consider a collection of v Regular Qualitative Features [RQlFs] $[Q1, Q2, ..., Qv]$ and only one single SQlF, Q^*. We thus have a total collection of $(v + 1)$ QlFs. A respondent is presented with a 'block' of questions of 'size' k+1 - typically a set of 'questionnaire' involving some k of the RSlFs and the SQlF and he/she is to provide only the Block Total Response [BTR] in terms of total 'score' [i.e., number of 'yes' responses] without divulging any separate information as to his/her status in respect of each of the $k + 1$ questions included in the questionnaire. It is believed that for any reasonable choice of k, the respondent is convinced about the honesty and integrity of the data collection agency and that, in its turn, will protect the privacy of the respondent and hence truthful response to Q^* will emerge! Since only information on the total score from a block is divulged by the respondent, it is termed Block Total Response Method/Technique [BTRM/BTRT]. In applications, there are a number of blocks [each of size $(k+1)$] of different sets of questionnaires - comprising of k RQlFs and the SQlF Q^* and, additionally, there is one more block containing all the v RQlFs. This is a must as, otherwise, unbiased estimation becomes impossible under the above framework. Basically, the SQlF acts as a 'Supplementary Question' and it is attached to every block of RQlFs, except the last one as mentioned above. The blocks to be used may be based on our notion of block designs such as Balanced Incomplete Block Designs [BIBDs].
The BTR procedure, as described by Raghavarao and Federer (1979), is likely to provide more 'perceived' protection and confidentiality to the respondents. The BTR approach is clear and concise to explain and adopt.

Further, the BTR procedure handles the case [of unknown proportion for the RQlF] in a natural manner and it also provides unbiased estimates of all the v RQlF-based proportions. This is a routine exercise involving linear models. The primary problem of estimation of P^* corresponding to the SQlF is discussed in the literature.

We should also mention in passing that while dealing with the SQlF Q^*, the respondent may use the RRT, if he/she so desires. Towards the beginning, Supplemented BIBDs were in use and methodologies were described and formulae were derived along these lines. These are also known in the design literature as Balanced Treatment Incomplete Block Designs [BTIBDs]. To reinforce the use of this fascinating technique, Nandy et al (2016) developed and discussed other variations of this method. These are briefly mentioned below.

3 Use of BIBD and Complimentary BIBD

As discussed in the literature, we start with a $BIBD[b, v, r, k, \lambda]$ which consists of b blocks, each of size k. The v treatments of the BIBD exclusively represent a set of v RQlFs so that each block contains exactly k of these RQlFs. We may then reinforce use of the RRT in the sense that in every block, we utilize the remaining $v - k$ RQlF's along with the SQlF, Q^*, in the formation of the RRT having $(v - k) + 1$ options for the respondent to choose from. As before, we associate a chance θ to Q^* and each of the rest in the RRT has associated with it the chance of $(1-\theta)/(v-k)$. The message is very clear and quite appealing. Each respondent has k clearly stated RQlF's to respond to and on the top of that, he/she has to respond to exactly one more question - chosen according to the RRT described above. Finally, he/she has to report only the total score based on $k+1$ QlFs and not any details! For example, with $v = 9$ and $k = 4$, the RRT will involve $4+1 = 5$ questions altogether and if $\theta = 0.20$, then the respondent has 20 per cent chance of picking up the SQlF and 16 per cent chance of picking up any one of the 5 RQlF's in the complementary block, besides the 4 RQlFs in the original block. Naturally, a total score of 3, for example, provided by a respondent in any of these b blocks, may arise in so many ways that it is virtually impossible to track down the actual ‘response category’ of the respondent(s) in respect of the SQlF, even if it is selected via the RRT!

To top it all, we need not start with a BIBD; any Binary Proper Equireplicate Block Design [BPEBD] with parameters $[b, v, r, k]$ would suffice. Of course, in all of the above, it is implicitly assumed that there is an additional block of all the v RSlFs for a group of respondents to attend to and provide the score in the form of BTR.

In Nandy et al (2016), the above framework was generalized in at least two directions. We do not enter into those technicalities. In this article, we start with a crucial observation instead.

4 Block Size (k+1) versus v?

In the entire literature on the use of BTRT, something undesirable and unnatural has crept in - without any serious objection to it! As we realize, we are banking on two sets of respondents : One set arises out of those in the b blocks wherein RQlFs and SQlF are, in a sense, blended together and each respondent provides a Block Total Response out of a collection of $(k+1)$ questions selected in the process. We may call this BLOCK TYPE I. The other set of respondents have to face all v RQlFs each. They may be said to form BLOCK TYPE II. In most cases it is quite plausible that $(k+1) << v$ and that would create a sense of uneasiness in the community of respondents belonging to Blocks of TYPE I and II. In other versions of the BTRT, this particular phenomenon of $(k+1)$ vs v remains unchanged and unnoticed as well. We propose to address this situation and suggest a remedy to ’close’ this gap!

Type I has BTR arising out of $(k+1)$ questions each and there are b such blocks. On the other hand, the only Block of Type II engages respondents with BTR based on v questions each. To bring equilibrium in the sense of allotting equal number of questions to all respondents, we may engage b^* Blocks of Type II which are formed as those of a BPEBD $(b^*, v, r^*, k^* = k+1)$. It is easy to construct such BPEBDs for given v and k. After data collection is over, it is a routine task to carry out the data analysis.

For Blocks of Type I(II), summing over the average response scores across all the $b(b^*)$ blocks and denoting it by Total Block Average (TBA-I(II)), we obtain :

$$E(TBA-I) = rT(P) + \theta P^* + [(1-\theta)/(v-k)](b-r)T(P).$$

On the other hand, for TBA-II, we may deduce

$$E(TBA-II) = r^*T(P).$$

From these two, we can deduce algebraic expression for (i) an unbiased estimate of P^* as also (ii) an expression for the variance of the estimate.

$$(i)\ [TBA-I]/\theta - [r + \frac{(1-\theta)(b-r)}{v-k}](TBA-II)/r^*\theta;$$

$$(ii)\ \sum_i \sigma_i^2(I)/n_I\theta^2 + [r + \frac{(1-\theta)(b-r)}{v-k}]^2 \sum_i \sigma_i^2(II)/n_{II}\theta^2 r^{*2}.$$

In the above, $\sigma_i^2(I)/n_I$ is the variance of the average response score for the ith block; $i = 1, 2, ..., b$. Likewise $\sigma_i^2(II)/n_{II}$ is the variance of the average response score for the ith block; $i = 1, 2, ..., b^*$. Moreover, because of repeated independent observations arising out of the respondents from each block, variance estimation [$\sigma_i^2(I)$ or $\sigma_i^2(II)$] is immediate.

Remark 1. In the supplementary part of Type I Blocks, we could adopt πPS design for selection of $t > 1$ QlFs, meeting $\pi = \pi^*$ for the SQlF. This we propose to study in the next section.

5 Block Size (k+t)

In the previous section, we have considered Blocks of Type I having BTR arising out of $(k+1)$ questions each and there are b such blocks. In each block we have k RQlFs and one additional QlF based on implementation of a selection rule - with a given probability of selection θ of the SQlF Q^*. In situations wherein $k << v$, the respondents might not be convinced enough about the extent of confidentiality attached to the above strategy. The number of questions to be handled by each respondent is $(k + 1)$ - the last one being selected through randomization and may turn out to be the SQlF itself. That's when lack of confidentiality may be suspected. To overcome this unpleasant situation, we propose to extend this strategy by referring to a selection of $t > 1$ QlFs out of the $(v - k) + 1$ QlFs. To make the selection rule simple, we may adopt Midzuno Sampling Scheme [Vide Hedayat and Sinha (1991)] and attach $\pi = \pi^* = t\theta$ to the SQlF. Following this scheme, one selects one unit [out of $(v - k) + 1$ units] with $P(SQlF) = p_0$ and $P(RQlF) = (1 - p_0)/(v - k)$ for each of the complimentary $(v-k)$ RQlFs in each block. The rest of the $(t-1)$ QlFs are selected as per SRSWOR sampling. The choice of p_0 is such that

$$t\theta = p_0 + [(1 - p_0)(t - 1)/(v - k)(v - k - 1)].$$

And this works provided $\theta > (v - k)^{-1}(v - k - 1)^{-1}$. Of course, for the Blocks of Part II, we consider a BPEBD with block size $k + t$ so that the blocks of both the types are comparable.

Remark 2. This approach gives more flexibility to the survey practitioners and increased confidence to the respondents - no matter which type of blocks they select for responding to the BTRs.

References

Chaudhuri, A. (2011). *Randomized Response and Indirect Questioning Techniques in Surveys*. CRC Press, Chapman and Hall, Taylor & Francis Group, FL, USA.

Chaudhuri, A. and Christofides, T.C. (2013). *Indirect Questioning in Sample Surveys*. Springer, Germany.

Chaudhuri, A. and Mukerjee, R.(1988). *Randomized Response : Theory and Applications*. Marcel and Dekker.

Hedayat, A.S. and Sinha, Bikas K. (1991) *Design and Inference in Finite Population sampling*. New York: Wiley.

Nandy, K., Marcovitz, M. and Sinha, B. K. (2015). Eliciting Information on Sensitive Features: Block Total Response Technique and Related Inference, 317 - 328. In *Handbook of Statistics*, Volume 34 : Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits.

Nayak, Tapan K. and Adeshiyan, S. A. (2009). A unified framework for analysis and comparison of randomized response surveys of binary characteristics. *Jour. Stat. Plann. & Inference*. **139**, 2757-2766.

Raghavarao, D., and Federer, W.T. (1979) Block Total Response as an Alternative to the Randomized Response Method in Surveys. *Journal of the Royal Statistical Society, Series B, Methodological*, **41**, 40 - 45.

Smith, L., Federer, W., and Raghavarao, D. (1974) A Comparison of Three Techniques for Eliciting Truthful Answers to Sensitive Questions. *Proceedings of the Social Statistics Association, American Statistical Association* pp. 447-452. Baltimore: American Statistical Association.

Warner, S.L. (1965) Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, **60**, 63-69.