A NOTE ON THE RANDOMIZED RESPONSE TECHNIQUE

Jagbir Singh, Temple University

1. Introduction

In surveys related to delicate questions, Warner (1965) introduced a randomized response technique for eliciting information, and thus estimating the proportion \( \pi \) of the population in the sensitive category \( A \). The technique consists in providing a spinner, with two outcomes \( A \) or not \( A \) with associated probabilities \( p \) and \( q = 1 - p \) to each respondent. The respondent spins the spinner unobserved by the interviewer and answers yes if he possesses the characteristic indicated by the pointer and no otherwise. It is obvious that the probability of a yes response, say \( \theta \), is

\[
\theta = p \pi + q(1 - \pi) = p + (2p - 1)\pi.
\]

If \( n_1 \) is the number of yes responses from a total of \( n \) interviews, and if all responses are truthful, Warner claimed that the maximum likelihood estimator (m.l.e.) of \( \theta \) is \( \hat{\theta} = n_1/n \) and consequently that of \( \pi \) is

\[
\hat{\pi} = \frac{n_1/n}{1 - (2p - 1)}/(2p - 1).
\]

provided \( p \neq 1/2 \). The purpose of this note is to point out that \( \hat{\theta} \), as well as \( \hat{\pi} \), are not the m.l.e.'s of the respective parameters. Indeed the m.l.e. of \( \theta \), say \( \bar{\theta} \), presented in this note is uniformly better than \( \hat{\theta} \) with respect to squared error loss. In other words, \( \hat{\theta} \) is not even admissible with respect to squared error loss. The same, of course, is true with \( \hat{\pi} \).

2. The Maximum Likelihood Estimator

Let us assume without any loss of generality that \( p > 1/2 \). It is of crucial significance to note that \( \theta \) is restricted to be in the interval \((p, q)\), since \( \pi \) is in \((0, 1)\). Thus the m.l.e. \( \bar{\theta} \) of \( \theta \) is

\[
\bar{\theta} = \begin{cases} 
 n_1/n & \text{if } p < n_1/n < q \\
 p & \text{if } p \geq n_1/n \\
 q & \text{if } n_1/n \geq q
\end{cases}
\]

Let \( b(x; n, \theta) \) be the binomial probability density function of \( n_1 \). We can write

\[
E(\bar{\theta}) = \theta + B
\]

where the bias \( B \) is

\[
B = \frac{1}{n} \sum_{x<np} (p-x)b(x; n, \theta) + \sum_{x=np}^{n} (np-x)b(x; n, \theta).
\]

The values of the first and the second summations are respectively positive and negative. However, overall magnitude and the sign of \( B \) will depend upon the values of the parameters \( n, p \), and \( \pi \). For fixed \( n \), \( p \), it can be shown that the bias evaluated at \( \pi \) is equal in magnitude to that at \( (1-\pi) \) but opposite in sign. We computed the bias magnitude for a limited choice of the parameters, and found it to be very small. It is easy to verify that \( \bar{\theta} \) is consistent.

To see that \( \bar{\theta} \) is uniformly better than \( \hat{\theta} \) with respect to squared error loss, we notice that:

\[
E(\bar{\theta} - \theta)^2 < E(\hat{\theta} - \theta)^2 = \sum_{x=np}^{n} \frac{(p-x)^2}{x^2np} b(x; n, \theta)
\]

Both the summations are negative. In the first summation \( x/n = p \). This means \( x/n - \theta = p - \theta < 0 \). Hence \( (p-x)^2 - (x/n - \theta)^2 < 0 \). This guarantees the value of the first summation to be negative. By the similar argument the second summation is also negative. Thus \( \bar{\theta} \) is not even admissible. The m.l.e. \( \bar{\pi} \) of \( \pi \) is obtained from \( \bar{\theta} \), since \( \pi \) is linearly related to \( \theta \).

\[
\bar{\pi} = \begin{cases} 
 \frac{n_1}{n} - (1-p)/2p - 1 & \text{if } p < n_1/n < p, p \neq 1/2 \\
 0 & \text{if } p \geq n_1/n \\
 1 & \text{if } n_1/n \geq p
\end{cases}
\]

Of course, statements which are true for \( \bar{\theta} \) relative to \( \hat{\theta} \) continue to be true for \( \bar{\pi} \) relative to \( \hat{\pi} \). Asymptotically, both \( \bar{\pi} \) and \( \bar{\pi} \) are equally efficient, but for small surveys \( \bar{\pi} \) is highly efficient. For example, if \( n = 10, p = 0.8 \), \( \bar{\pi} = 1/12 \), then the relative efficiency of \( \bar{\pi} \) is 1.62.

In conclusion, it should be said that similar observations can be made for the follow-up research on this topic.

3. Summary

In surveys related to delicate questions, Warner (Journal of the American Statistical Association, 60 (1965), 63-69) introduced a randomized response technique for eliciting information, and thus estimating the proportion \( \pi \) of the population in the sensitive category. The estimator of \( \pi \) obtained was called, not only by Warner but also subsequently in the literature, as the maximum likelihood estimator. It is shown here that Warner's estimator is indeed neither the maximum likelihood estimator nor even admissible. The maximum likelihood estimator given in this note is uniformly better than that of Warner's with respect to squared error loss.

References