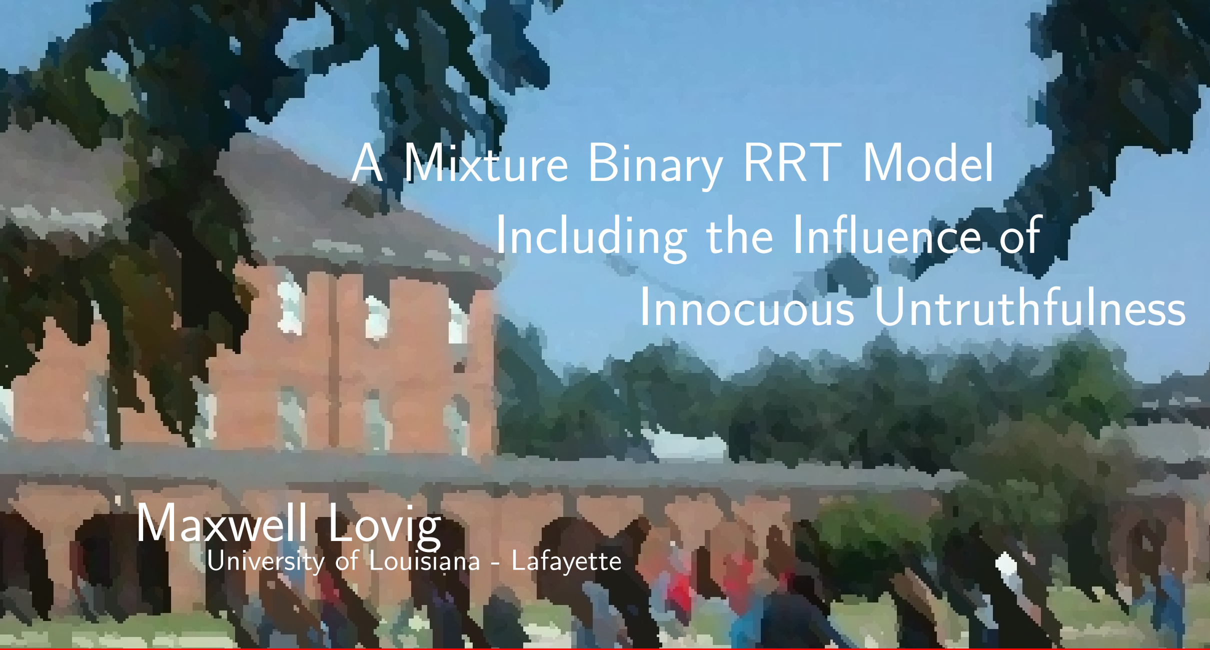


A Mixture Binary RRT Model Including the Influence of Innocuous Untruthfulness

Maxwell Lovig
University of Louisiana - Lafayette

Introduction

The Innate Responsibility for Survey Making and Taking

When creating a survey there is a subtle sense of cooperation between two parties: the survey maker and the survey taker. The former makes the rules and the latter decides when to follow them.

The survey maker wants information of its demographic

The survey taker wants their information to be private and secure

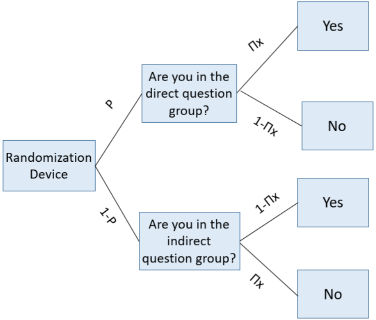
How does RRT accomplish this

Binary RRT (Randomized Response Techniques) effectively randomly changes the yes/no response from a survey taker unbeknownst to the survey maker. This is usually accomplished through a secondary question which occurs at a random rate of 1-p. With the sensitive question (with a true population occurrence at π_x) at probability p.

Sensitive Question:
"Have you purposefully neglected care of your parents?"

Alternate Questions:
"Have you never purposefully neglected care of your parents?" (Warner's Model)
-OR-
"Is your favorite color blue?
(Greenberg's Model)

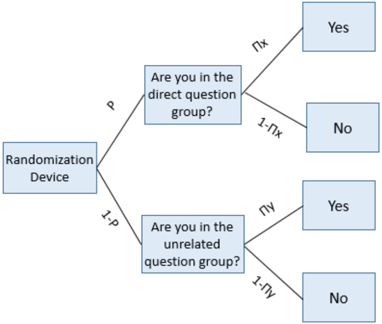
Indirect Question Model [Warner 1965]



$$P_y = p\pi_x + (1-p)(1-\pi_x)$$

$$\widehat{\pi_x} = \frac{\widehat{P_y} - (1-p)}{2p-1}; p \neq \frac{1}{2}$$

Unrelated Question Model [Greenberg et al. 1969]



$$P_y = p\pi_x + (1-p)\pi_y$$

$$\widehat{\pi_x} = \frac{\widehat{P_y} - (1-p)\pi_y}{p}$$

$$P_y(1-P_y)$$

Untruthfulness even under RRT

Now imagine yourself as a participant of a survey. Of course there is some level of privacy in which you would answer truthfully for. If the survey maker does not provide this level of privacy (i.e. a value of p which is too high) you may still find lying is a viable option and as such RRT will usually never have 100% compliance to instructions.

For example given the following with Greenberg's Model parameters:

$$n = 500,\ \pi_x = .3,\ \pi_y = .1$$

p	1-p	A	$\hat{\pi}$	$\widehat{MSE}$	MSE
.5	.5	1	.29984	.00128	.00128
.5	.5	.9	.27023	.00209	.00218
.5	.5	.8	.23967	.00478	.00488
.7	.3	1	.30023	.00076	.00075
.7	.3	.9	.27006	.00160	.00165
.7	.3	.8	.24035	.00424	.00435

Solution to remaining untruthfulness

Proposed by Young, Gupta, and Parks 2019 a primer question was proposed to ask the participant "Do you trust this model" (i.e. would you respond truthfully). As this information itself is sensitive it must also be placed in a RRT survey format.

This extra information gives us a value of A. The proportion of participants who respond truthfully to the instructions of RRT and would not lie with the current parameters in RRT.

Part 1: Mixture Model

The conundrum

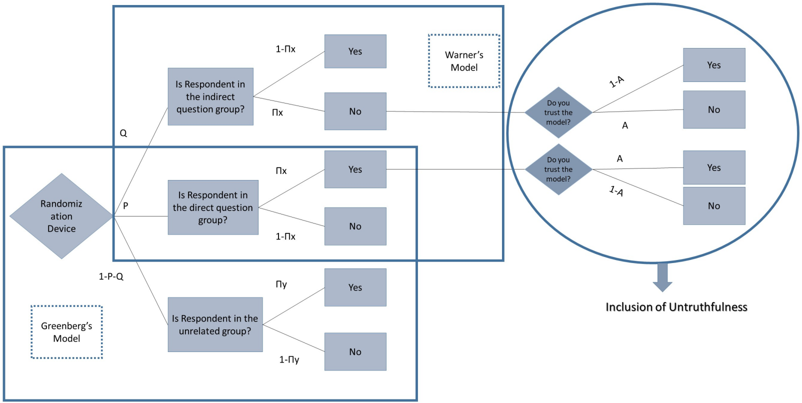
Efficiency: The ability for the researcher to get close to the true proportion of π_x, measured by MSE

Privacy: The sense of security in the participant of the survey. This have many diverse measures.

Overall, Greenberg's model tends to have better efficiency measures while Warner's model tends to have better privacy.

So it seems that a "mixture" of the two models would be able to strike a balance in these two categories.

Proposed Mixture Model



The Proposed Mixture Model Under Untruthfulness

Derivation of MSE When Estimating for A and π_x

$p_0 =$ The proportion of the direct question used in a Greenberg model to estimate truthfulness.
$\pi_{y0} =$ The proportion of people who would answer yes to the unrelated question in the Greenberg model to estimate A
p, q, π_x, π_y, A are all the same from the introduction of the mixture model

Question 1: (With Greenberg) Do you trust the model?

$$P_{y0} = P_0(Yes) = p_0 A + (1 - p_0)\pi_{y0}$$

$$\hat{A} = \frac{\hat{P}_{y0} - (1 - p_0)\pi_{y0}}{p_0} \quad (1)$$

$$E(\hat{A}) = A, \ Var(\hat{A}) = \frac{P_{y0}(1 - P_{y0})}{(n-1)p_0^2} \quad (2)$$

Question 2: (With Mixture Model) Do you have sensitive trait?

$$P_y = P(Yes) = \pi_x A(p-q) + q + (1-p-q)\pi_y$$

$$\hat{\pi}_x = \frac{\hat{P}_y - q - (1-p-q)\pi_y}{\hat{A}(p-q)};\ p \neq q \tag{3}$$

$$E(\hat{P}_y) = P_y = \pi_x A(p-q) + q + (1-p-q)\pi_y,\ Var(\hat{P}_y) = \frac{P_y(1-P_y)}{(n-1)(p-q)^2};\ p \neq q \tag{4}$$

Then using first order Taylor's approximation, $\hat{\pi}_x$ can be approximated by:

$$\hat{\pi}_x \approx \frac{P_y - q - (1-p-q)\pi_y}{A(p-q)} - (\hat{A} - A)\left[\frac{P_y - q - (1-p-q)\pi_y}{A^2(p-q)}\right] + (\hat{P}_y - P_y)\left[\frac{1}{A(p-q)}\right]] \tag{5}$$

$$[where\ p \neq q]$$

It is easy to see from (5) that,

$$E(\hat{\pi}_x) = \pi_x$$

$$Var(\hat{\pi}_x) = \left[\frac{P_y - q - (1-p-q)\pi_y}{A^2(p-q)}\right]^2 Var(\hat{A}) + \left[\frac{1}{A(p-q)}\right]^2 Var(\hat{P}_y)$$

$Var(\hat{A})$ is given in (2) and $Var(\hat{P}_y)$ is given in (4)

Privacy Loss

Lanke 1976, in previous work, described privacy loss as,

$$\delta = \text{Max}(\eta_1, \eta_2)$$

$$\eta_1 = Pr(S|Y) = \frac{pA\pi_x + (1-p-q)\pi_x\pi_y + q\pi_x(1-A)}{\pi_x A(p-q) + q + (1-p-q)\pi_y}$$

$$\eta_2 = Pr(S|N) = \frac{qA\pi_x + (1-p-q)(1-\pi_y)\pi_x + p\pi_x(1-A)}{1 - [\pi_x A(p-q) + q + (1-p-q)\pi_y]}$$

Results

Result: δ is at its minimum if p $=$ q, $\min(\delta) = \pi_x$

Proof:

$\implies$ Let $p = q$, then

$$P(S|Y) = \frac{\pi_x[pA + (1-p-q)\pi_y + q(1-A)]}{\pi_x A(p-q) + q + (1-p-q)\pi_x} = \pi_x$$

$$P(S|N) = \frac{\pi_x[qA + (1-p-q)(1-\pi_y) + p(1-A)]}{\pi_x A(q-p) + p + (1-p-q)(1-\pi_y)} = \pi_x$$

Hence $\delta = \pi_x, \ \textit{when } p = q$

Now Assume,

$$
\begin{aligned}
&\exists p \text{ and } \exists q \text{ st. } P(S|Y) < \pi_x \text{ and } P(S|N) < \pi_x \\
&\implies \frac{P(S \cap Y)}{P(Y)} < \pi_x \text{ and } \frac{P(S \cap N)}{P(N)} < \pi_x \\
&\implies P(S \cap Y) < \pi_x P(Y) \text{ and } P(S \cap N) < \pi_x P(N) \\
&\implies P(S \cap Y) + P(S \cap N) < \pi_x P(Y) + \pi_x P(N) \\
&\implies P(S) < \pi_x
\end{aligned}
$$

This creates a contradiction proving that $Min\ \delta = \pi_x$ and it is attained at $p = q$

Primary Protection

If one wants to transform privacy loss into Primary Protection then the following formula can be applied. This measure was introduced by Fligner, Policello, and Singh 1977

$$PP = \frac{1-\delta}{1-\pi_x};$$

$$\delta \in [\pi_x, 1] \textit{ and } \pi_x \in [0, 1]$$

Unified Measure of Privacy and Efficiency

Unified Measure of Privacy and Efficiency

Researchers have to deal with two factors for a RRT model. One factor is the efficiency of the model and another is the Privacy.Gupta et al. 2018came up with following Unified Measure of Privacy and Efficiency for RRT model.

$$M = \frac{Var(\hat{\mu_x})}{PL};\ PL = E(Z-Y)^2$$

Proposed Unified Measure

From the concepts of Lanke 1976 loss to privacy, Fligner, Policello, and Singh 1977 Primary Protection and Gupta et al. 2018 Unified Measure of Privacy and Efficiency, we propose the following Unified Measure,

$$\mathbb{M} = \frac{PP}{MSE}$$

Simulations: $\pi_x = .4, \pi_y = .1, p_0 = .7, \pi_{y0} = .1.$

Subscript Meanings:
S - simulated values, T - theoretical values
1 - Accounting A, 2 - Not Accounting A
Rose is Warner's. Blue is Greenberg's, No color is Mixture
Bolded Values are theoretical, non-bolded are empirical

p	q	1-p-q	A	$\hat{\pi}_1$	MSE_1	$\hat{\pi}_2$	MSE_2	PP	M_1	M_2
.4	0	.6	1	.3998	.0023	.3994	.0021	.2726	12.2388	128.1228
					.0023		**.0022**	**.2727**	**119.7078**	**126.8913**
.4	0	.6	.8	.4006	.0031	.3103	.0083	.3190	102.6930	38.6041
					.0032		**.0085**	**.3191**	**98.6590**	**37.3304**
.4	.1	.5	1	.4007	.0044	.4003	.0043	.5556	125.1477	129.8006
					.0045		**.0044**	**.5555**	**122.9716**	**126.5855**
.4	.1	.5	.8	.3997	.0066	.3190	.0107	.6101	92.3568	56.8758
					.0067		**.0108**	**.6097**	**9.9995**	**56.5176**
.4	.6	0	1	.4012	.01275	.4006	.0126	.8331	65.3209	66.3605
					.0126		**.0125**	**.8333**	**65.9596**	**66.6400**
.4	.6	0	.8	.3997	.0193	.3193	.0187	.8617	44.6630	46.0447
					.0197		**.0189**	**.8621**	**43.7237**	**45.6000**
.7	0	.3	1	.3998	.0010	.3997	.0009	.0969	97.0556	11.8450
					.0010		**.0009**	**.0968**	**96.4091**	**11.6230**
.7	0	.3	.8	.4012	.0013	.3206	.0071	.1177	89.9563	16.6216
					.0015		**.0073**	**.1181**	**8.9974**	**16.2355**
.7	.1	.2	1	.4002	.0014	.3999	.0013	.3333	236.6728	258.1805
					.0014		**.0013**	**.3333**	**236.1485**	**259.8958**
.7	.1	.2	.8	.4003	.0020	.3201	.0076	.3850	195.2027	5.6466
					.0021		**.0077**	**.3855**	**181.9029**	**5.0634**
.7	.3	0	1	.4003	.0032	.3999	.0031	.6523	203.5105	211.5794
					.0032		**.0031**	**.6522**	**201.2790**	**209.6190**
.7	.3	0	.8	.3997	.0049	.3194	.0096	.7013	143.3820	73.2478
					.0050		**.0095**	**.7009**	**139.1164**	**73.6955**

Part 2: Innocuous untruthfulness

Back to Untruthfulness

This issue is prevalent in the estimation of P_y where the value of $A < 1$ can impact the the variance but with a large enough n, the variance is a nonissue. What is a larger issue to the model is the bias under untruthfulness. For all 3 models the bias is the same, with A as the proportion of respondents who will answer the sensitive question honestly.

$$Bias(\hat{\pi}_x) = \pi_x(A - 1)$$

Avoiding Untruthfulness has been the purpose of RRT models since first introduced [Warner 1965]. The cooperation of participants has been further elaborated on with the implementation of the innocuous question model [Greenberg et al. 1969].

Another Type of Untruthfulness

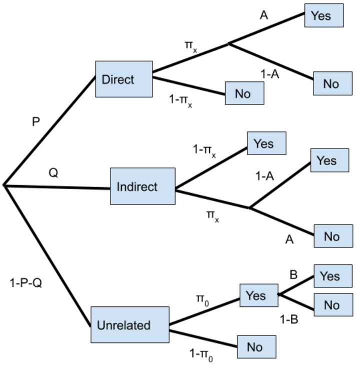
The innocuous question can also cause untruthfulness. For very social sensitive topics, respondent may not respond yes to a question regardless of which question they are given. Due to the response of a "yes" implicating the participant even though they aren't asked the sensitive question

This leads to issues in the design of RRT Models where a innocuous question was meant to be a safe area. As such we must consider which models are more robust to these types of untruthfulness [Edgell, Himmelfarb, and Duchan 1982].

TABLE 1
Questions and Percentages Who Did Not Follow the Device

Question	Device Directed Answer	Percentage Who Did Not Answer as Directed
1. Have you ever tried to make an elderly person feel less lonely by visiting with that person?	No	4%
2. Do you enjoy reading soft-core pornography (e.g., Playboy, Playgirl)?	No	4%
3. Have you ever had complete sexual intercourse?	No	6%
4. Do you approve of premarital sex for engaged couples?	No	6%
5. Have you ever falsified your income tax report?	Yes	15%
6 Are you in favor of building a nuclear power plant at Marble Hill?	Yes	11%
7. Have you ever had a homosexual experience?	Yes	26%

Effect of innocuous untruthfulness



Flowchart of The Mixture Model with the impact of innocuous untruthfulness

Assume A and B are independent, and as we want the answer "yes" to be the sensitive option we also assume $p > .5$.

Effect on Estimation

Consider the standard estimator $\hat{\pi}_x$,

$$\implies Bias(\hat{\pi}_x) = \pi_x(A-1) + \pi_0(B-1)\left(\frac{1-p-q}{p-q}\right)$$

Meaning the Bias for Warner, and Greenberg's model is

$$Bias(\hat{\pi}_{x_W}) = \pi_x(A-1), \quad Bias(\hat{\pi}_{x_G}) = \pi_x(A-1) + \pi_0(B-1)\left(\frac{1-p}{p}\right)$$

The variance of the model is unchanged from before, besides the adjustment to P_y

$$var(\hat{\pi}_x) = \frac{P_y(1-P_y)}{n(p-q)^2}, \text{ where } P_y = p\pi_x A + q\pi_x(1-A) + q(1-\pi_x) + (1-p-q)\pi_0 B$$

Effect on Estimation (Cont).

Leading to a MSE of

$$\implies MSE(\hat{\pi}_x) = \frac{P_y(1-P_y)}{n(p-q)^2} + \left(\pi_x(A-1) + \pi_0(B-1)\left(\frac{1-p-q}{p-q}\right)\right)^2$$

Where $q = 0$ leads to Greenberg's Model and $q = 1 - p$ leads to Warner's model. When $B < 1$ it is unclear which model leads to a lower MSE. Greenberg's model has a more desirable variance where as Warner's model has the more desirable bias. The Mixture model is the tool we can use to find a model such that the effect of A & B are at a minimum

If we have some indicator of the values for A and B, we are able to find a q which can lead to a good estimator. We can solve for the q which minimizes MSE, assuming n is sufficiently large such that $Bias(\hat{\pi}_x) >> (\hat{\pi}_x)$ with the following equation

$$\frac{\partial}{\partial q} MSE(\hat{\pi}_x) \approx \frac{\partial}{\partial q} Bias(\hat{\pi}_x)^2 = \frac{\partial}{\partial q} \left(\pi_x(A-1) + \pi_0(B-1) \left(\frac{1-p-q}{p-q} \right) \right)^2 = 0$$

Which occurs at

$$q \approx \frac{\pi_x p(A-1) - \pi_0(p-1)(B-1)}{\pi_x(A-1) + \pi_0(B-1)}, \text{ assuming q is a feasible solution}$$

Lacking Estimation

There is a method introduced which is able to estimate for the value of A by asking the participants if they trusted the model Young, Gupta, and Parks 2019. As this question is of itself sensitive it was put through a RRT model of its own. There are two concerns with using this approach for our needs.

Firstly, the ability for someone who doesn't trust the RRT process to answer truthfully for the estimate of A is unknown and perhaps not likely to occur.

Secondly, the proportion of participants who would answer this question, assuming perfect truthfulness would be estimating $P(A \bigcap B) = (A)(B)$. The actual values of A and B are unrecoverable and as such this method is not reasonable for our uses.

Robustness of Basic RRT models under untruthfulness

We present the following metric, R, for the robustness of RRT models to untruthfulness of both the sensitive and innocuous questions. Let the possible value for $A \in [a_A, b_A]$ and for $B \in [a_B, b_B]$, given π_x, π_0, q, and n, we have

$$R = \int_{a_A}^{b_A} \int_{a_B}^{b_B} MSE(\hat{\pi}_x) \, dB \, dA$$

For the rest of this paper we have $a_A = b_B = .8$ and $b_A = b_B = 1$.

Tables of Optimal Models

We vary the value of $p > .5$, n and π_x, π_0 to find the q which give us the lowest value of R (q is rounded to 2 significant figures). The value of R was calculated through numerical integration in Julia.

		π_x				
		.05	.10	.15	.20	.25
	.5	.31	.31	.31	.30	.30
	.6	.24	.24	.23	.23	.22
p	.7	.17	.17	.16	.15	.15
	.8	.11	.10	.09	.09	.08
	.9	.05	.04	.03	.03	.02

		π_x				
		.05	.10	.15	.20	.25
	.5	.26	.25	.25	.25	.25
	.6	.22	.22	.21	.21	.21
p	.7	.16	.16	.15	.15	.14
	.8	.10	.10	.09	.08	.08
	.9	.04	.04	.03	.03	.02

Values for q which minimize R for mixture RRT with $n = 1000$ (left) and $n = 100$ (right), $\pi_0 = .15$

What To Take Away

The value of q which minimizes the $MSE(\hat{\pi}_x)$ belongs to a mixture model for all parameters we tested.

Meaning that if we consider a uniform level of untruthfulness with the sensitive and innocuous question from 0.8 to 1 then choose the mixture model.

It is also worth nothing that the ideal q does not vary that much given changes in π_x and n.

The only decisive choice they are left to make is the choice in p which can hopefully keep the level of untruthfulness relatively low.

Future Work

Firstly, there should be further investigation into how correlated the values of A and B are. This would validate or disprove the assumption A and B are independent.

Secondly, there is yet to be any studies showing how the values of p and q control untruthfulness of all kinds. Knowing this information will help alleviate most of the guesswork in creating and RRT model.

Unfortunately, both of these questions require empirical results in the field which have yet to be conducted.

References (1)

- Edgell, Stephen E., Samuel Himmelfarb, and Karen L. Duchan (1982). 'Validity of Forced Responses in a Randomized Response Model'. In: **Sociological Methods Research** 11.1, pp. 89–100.
- Fligner, Michael A., George E. Policello, and Jagbir Singh (1977). 'A comparison of two randomized response survey methods with consideration for the level of respondent protection'. In: **Communications in Statistics - Theory and Methods** 6.15, pp. 1511–1524.
- Greenberg, Bernard G. et al. (1969). 'The Unrelated Question Randomized Response Model: Theoretical Framework'. In: **Journal of the American Statistical Association** 64.326, pp. 520–539.
- Gupta, Sat et al. (2018). 'A Unified Measure of Respondent Privacy and Model Efficiency in Quantitative RRT Models'. In: **Journal of Statistical Theory and Practice** 12.3, pp. 506–511.

References (2)

- Lanke, Jan (1976). 'On the Degree of Protection in Randomized Interviews'. In: **International Statistical Review / Revue Internationale de Statistique** 44.2, pp. 197–203.
- Warner, Stanley L. (1965). 'Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias'. In: **Journal of the American Statistical Association** 60.309, pp. 63–69.
- Young, Amber, Sat Gupta, and Ryan Parks (2019). 'A binary unrelated-question RRT model accounting for untruthful responding'. In: **Involve** 12.7, pp. 1163–1173.