Tackling Malicious Behaviour in Randomised Response Techniques

Palash Sarkar

Applied Statistics Unit
Indian Statistical Institute, Kolkata
India
palash@isical.ac.in

Advanced Survey Research Centre,
11th December 2011

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Are you using LPG cylinders illegally?

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Are you using LPG cylinders illegally?

Have you taken bribes in the past?

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Are you using LPG cylinders illegally?

Have you taken bribes in the past?

Have you avoided paying income tax?

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Are you using LPG cylinders illegally?

Have you taken bribes in the past?

Have you avoided paying income tax?

Do you have communal feelings?

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Are you using LPG cylinders illegally?

Have you taken bribes in the past?

Have you avoided paying income tax?

Do you have communal feelings?

Do you believe in the supremacy of upper castes?

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Are you using LPG cylinders illegally?

Have you taken bribes in the past?

Have you avoided paying income tax?

Do you have communal feelings?

Do you believe in the supremacy of upper castes?

Are you a Maoist sympathiser?

Some Sensitive Socio-Political Questions

Do you use ‘hooking’ for (illegal) electric connection?

Are you using LPG cylinders illegally?

Have you taken bribes in the past?

Have you avoided paying income tax?

Do you have communal feelings?

Do you believe in the supremacy of upper castes?

Are you a Maoist sympathiser?

Caveat: Some things are possibly best left unknown.

Survey on Sensitive Questions

All of these questions have Boolean (yes/no) answers.

Underlying population is divided into yes/no groups.

The purpose of a survey would be to estimate the yes/no proportions.

Survey on Sensitive Questions

All of these questions have Boolean (yes/no) answers.

Underlying population is divided into yes/no groups.

The purpose of a survey would be to estimate the yes/no proportions.

Survey respondents would be accordingly divided into two groups.

For each question, one group will almost always answer correctly, while the other group will almost always either not answer or answer incorrectly.

- Fear of been stigmatised.
- Fear of been persecuted.

Randomised Response Technique: Warner (1965)

A simple formulation: Toss a coin; answer ‘yes’ if it comes up tails; answer truthfully if it comes up heads.

Randomised Response Technique: Warner (1965)

A simple formulation: Toss a coin; answer ‘yes’ if it comes up tails; answer truthfully if it comes up heads.

Consider the ‘no’-group among the respondents; about half of them will get tails and answer ‘yes’; the other half will answer truthfully and say ‘no’.

Randomised Response Technique: Warner (1965)

A simple formulation: Toss a coin; answer ‘yes’ if it comes up tails; answer truthfully if it comes up heads.

Consider the ‘no’-group among the respondents; about half of them will get tails and answer ‘yes’; the other half will answer truthfully and say ‘no’.

So, if $x\%$ of the answers are ‘no’, then among the respondents about $2x\%$ are actually ‘no’.

Randomised Response Technique: Warner (1965)

A simple formulation: Toss a coin; answer ‘yes’ if it comes up tails; answer truthfully if it comes up heads.

Consider the ‘no’-group among the respondents; about half of them will get tails and answer ‘yes’; the other half will answer truthfully and say ‘no’.

So, if $x\%$ of the answers are ‘no’, then among the respondents about $2x\%$ are actually ‘no’.

The interviewer gets the desired estimate.

Randomised Response Technique: Warner (1965)

A simple formulation: Toss a coin; answer ‘yes’ if it comes up tails; answer truthfully if it comes up heads.

Consider the ‘no’-group among the respondents; about half of them will get tails and answer ‘yes’; the other half will answer truthfully and say ‘no’.

So, if $x\%$ of the answers are ‘no’, then among the respondents about $2x\%$ are actually ‘no’.

The interviewer gets the desired estimate.

Respondents answering ‘yes’ get plausible deniability: a respondent can claim that the coin came up tails.

Warner's Formulation

Warner's original formulation: a responder tells the truth with some fixed pre-determined probability $p > 1/2$ and lies with probability $(1 - p)$.

Eg.: roll a die; lie if it comes up 1 or 2; otherwise tell the truth.

Warner's Formulation

Warner's original formulation: a responder tells the truth with some fixed pre-determined probability $p > 1/2$ and lies with probability $(1 - p)$.

Eg.: roll a die; lie if it comes up 1 or 2; otherwise tell the truth.

Suppose among n respondents, x is the size of the 'yes'-group and R is the random variable counting the actual number of 'yes' responses.

Warner's Formulation

Warner's original formulation: a responder tells the truth with some fixed pre-determined probability $p > 1/2$ and lies with probability $(1 - p)$.

Eg.: roll a die; lie if it comes up 1 or 2; otherwise tell the truth.

Suppose among n respondents, x is the size of the 'yes'-group and R is the random variable counting the actual number of 'yes' responses.

Then $E(R) = px + (1 - p)(n - x)$ and so

$$x = \frac{E(R) - n(1 - p)}{2p - 1}.$$

Warner's Formulation

Warner's original formulation: a responder tells the truth with some fixed pre-determined probability $p > 1/2$ and lies with probability $(1 - p)$.

Eg.: roll a die; lie if it comes up 1 or 2; otherwise tell the truth.

Suppose among n respondents, x is the size of the 'yes'-group and R is the random variable counting the actual number of 'yes' responses.

Then $E(R) = px + (1 - p)(n - x)$ and so

$$x = \frac{E(R) - n(1 - p)}{2p - 1}.$$

R is an estimate for $E(R)$; given R, we can estimate x.

Other Formulations

Innocuous question: truthfully answer the question with probability p and with probability $(1 - p)$ answer an unrelated innocuous question such as ‘flip a coin: did you get tails?’.

Other Formulations

Innocuous question: truthfully answer the question with probability p and with probability $(1 - p)$ answer an unrelated innocuous question such as ‘flip a coin: did you get tails?’.

Polychotomous RRT: A respondent can belong to one of m mutually exclusive groups some of which are stigmatising. Suppose π_i is the proportion of people who belong to group i.

Other Formulations

Innocuous question: truthfully answer the question with probability p and with probability $(1 - p)$ answer an unrelated innocuous question such as ‘flip a coin: did you get tails?’.

Polychotomous RRT: A respondent can belong to one of m mutually exclusive groups some of which are stigmatising. Suppose π_i is the proportion of people who belong to group i.

Fix probabilities p_0 and $p_1, \ldots, p_m$, such that $p_0 + p_1 + \cdots + p_m = 1$; a respondent either reveals her true type t with probability p_0; or answers i with probability p_i.

Other Formulations

Innocuous question: truthfully answer the question with probability p and with probability $(1 - p)$ answer an unrelated innocuous question such as 'flip a coin: did you get tails?'.

Polychotomous RRT: A respondent can belong to one of m mutually exclusive groups some of which are stigmatising. Suppose π_i is the proportion of people who belong to group i.

Fix probabilities p_0 and $p_1, \ldots, p_m$, such that $p_0 + p_1 + \cdots + p_m = 1$; a respondent either reveals her true type t with probability p_0; or answers i with probability p_i. Let q_i be the proportion of respondents who answer i; then $q_i = p_0\pi_i + p_i$ and so $\pi_i = (q_i - p_i)/p_0$.

Other Formulations

Innocuous question: truthfully answer the question with probability p and with probability $(1-p)$ answer an unrelated innocuous question such as 'flip a coin: did you get tails?'.

Polychotomous RRT: A respondent can belong to one of m mutually exclusive groups some of which are stigmatising. Suppose π_i is the proportion of people who belong to group i.

Fix probabilities p_0 and $p_1, \ldots, p_m$, such that $p_0 + p_1 + \cdots + p_m = 1$; a respondent either reveals her true type t with probability p_0; or answers i with probability p_i. Let q_i be the proportion of respondents who answer i; then $q_i = p_0\pi_i + p_i$ and so $\pi_i = (q_i - p_i)/p_0$.

Arijit Chaudhuri and Rahul Mukherjee. Randomized Response: Theory and Techniques, Marcel Dekker, Inc., 1988,

Malicious Intentions

Consider the following question:

> Do you believe in the supremacy of the upper castes?

Malicious Intentions

Consider the following question:

> Do you believe in the supremacy of the upper castes?

Those who do have such belief may wish to bias the outcome of the survey by *always* answering truthfully.

Malicious Intentions

Consider the following question:

> Do you believe in the supremacy of the upper castes?

Those who do have such belief may wish to bias the outcome of the survey by *always* answering truthfully.

Those who do not have such belief will (probably) follow the protocol honestly.

Malicious Intentions

Consider the following question:

Do you believe in the supremacy of the upper castes?

Those who do have such belief may wish to bias the outcome of the survey by *always* answering truthfully.

Those who do not have such belief will (probably) follow the protocol honestly.

The outcome of the survey will be biased.

Usual randomised response techniques do not provide protection against such behaviour.

(Cryptographic) RRT Protocol

Two-party protocol: Interviewer and Responder.

(Cryptographic) RRT Protocol

Two-party protocol: Interviewer and Responder.

The responder has a private bit b.

Both parties perform some private computations.

The parties exchange some information.

Computations and communications can be interleaved.

(Cryptographic) RRT Protocol

Two-party protocol: Interviewer and Responder.

The responder has a private bit b.

Both parties perform some private computations.

The parties exchange some information.

Computations and communications can be interleaved.

At the end of the protocol, the interviewer gets to know a bit c which is equal to b with probability p and equal to $1 - c$ with probability $1 - p$.

CRRT Protocol: Desired Properties

Plausible Deniability: The responder should be able to claim with reasonable probability that the bit learned by the interviewer is not b.

CRRT Protocol: Desired Properties

Plausible Deniability: The responder should be able to claim with reasonable probability that the bit learned by the interviewer is not b.

Bounded Bias: The probability that $c = b$ is at most p even if the responder maliciously deviates from the protocol.

- Ensures that the responder cannot bias the outcome of the protocol in any way other than changing its own secret bit.

Ideal World Execution

There is a trusted third party (TTP).

The responder sends the secret bit b to TTP.

The TTP generates a bit c such that $c = 0$ with probability p ($p > 1/2$).

The TTP provides the interviewer with the bit $b \oplus c$.

Ideal World Execution

There is a trusted third party (TTP).

The responder sends the secret bit b to TTP.

The TTP generates a bit c such that $c = 0$ with probability p ($p > 1/2$).

The TTP provides the interviewer with the bit $b \oplus c$.

If the responder is honest, then the interviewer gets b with probability p.

Ideal World Execution: Security

Plausible deniability: The responder can always claim that the secret bit was $1 \oplus b$.

Ideal World Execution: Security

Plausible deniability: The responder can always claim that the secret bit was $1 \oplus b$.

Bounded bias:

- Suppose the responder sends a bit y to TTP which is equal to b with probability q.
- Then the TTP sends the bit $z = c \oplus y$ to the interviewer.
  $\Pr[z = b] = qp + (1 - q)(1 - p)$;
  this probability is between $(1 - p)$ and p.

Real World: The World is Not Ideal

The goal of a real world protocol is to simulate the ideal world protocol without a TTP.

Real World: The World is Not Ideal

The goal of a real world protocol is to simulate the ideal world protocol without a TTP.

Behaviour of the parties.

- Honest-but-curious: follows the protocol honestly.
- Malicious: deviates from the protocol.

Real World CRRT Parties

- **Interviewer:**
  - Honest-but-curious: follows the protocol, but may perform computation to learn the secret bit of the responder.
  - Malicious: can deviate from the protocol in an attempt to learn the secret bit of the responder.

Real World CRRT Parties

- **Interviewer:**
  - Honest-but-curious: follows the protocol, but may perform computation to learn the secret bit of the responder.
  - Malicious: can deviate from the protocol in an attempt to learn the secret bit of the responder.
- **Responder:**
  - Honest-but-curious: follows the protocol but is interested in learning the outcome of the survey.
  - Malicious: can deviate from the protocol in an attempt to bias the outcome of the survey.

CRRT Schemes and Related Work

H. Kikuchi, J. Akiyama, G. Nakamura and H. Gobioff. Stochastic Voting Protocol to Protect Voters Privacy. In 1999 IEEE Workshop on Internet Applications.

- 'Rediscovered' the notion of RRT.
- Possibly the first paper to deal with the problem of malicious bias.

CRRT Schemes and Related Work

H. Kikuchi, J. Akiyama, G. Nakamura and H. Gobioff. Stochastic Voting Protocol to Protect Voters Privacy. In 1999 IEEE Workshop on Internet Applications.

- 'Rediscovered' the notion of RRT.
- Possibly the first paper to deal with the problem of malicious bias.

A. Ambainis, M. Jakobsson and H. Lipmaa. Cryptographic Randomized Response Techniques. In Public Key Cryptography, 2004.

- The first paper which describes CRRT as an enhancement of RRT.
- Provides formal definitions of security requirments.
- Describes protocols achieving both plausible deniability and bounded bias.
- The protocols are quite complex and combine several sophisticated cryptographic primitives.

CRRT Schemes and Related Work

T. Moran and M. Naor. Polling With Physical Envelopes: A Rigorous Analysis of a Human-Centric Protocol. In Eurocrypt 2006.

- Provides the ideal/real world formulation of CRRT security.
- Describes two protocols which can be carried out without using a computer.
  - The first protocol assumes that the responder is honest and protects against malicious interviewer.
  - The second protocol assumes the the interviewer is honest and protects against malicious responder.
- Describes a rather impractical protocol which is secure against both malicious responder and malicious interviewer.

CRRT Schemes and Related Work

P. Golle, F. McSherry and I. Mironov. Data Collection With Self-Enforcing Privacy. ACM Transactions on Information Systems and Security, 2008.

- Introduces the notion of a ‘privacy bond’ to tackle the problem of trustworthiness of the interviewer.

CRRT Schemes and Related Work

P. Golle, F. McSherry and I. Mironov. Data Collection With Self-Enforcing Privacy. ACM Transactions on Information Systems and Security, 2008.

- Introduces the notion of a ‘privacy bond’ to tackle the problem of trustworthiness of the interviewer.

M. Z. Ashrafi and S. K. Ng. Collusion-Resistant Anonymous Data Collection Method. ACM Conference on Knowledge Discovery and Data Mining, 2009.

- Considers the problem of collusion attack on a respondent involving the interviewer and peer respondents.

Moran-Naor Physical CRRT Protocol-1

Requires scratch-off cards.

Moran-Naor Physical CRRT Protocol-1

Requires scratch-off cards.

Interviewer prepares card:

- A card has four scratchable bubbles arranged in two rows of two bubbles each.
- In each row, the bubbles hide the words ‘yes’ and ‘no’.

Moran-Naor Physical CRRT Protocol-1

Requires scratch-off cards.

Interviewer prepares card:

- A card has four scratchable bubbles arranged in two rows of two bubbles each.
- In each row, the bubbles hide the words ‘yes’ and ‘no’.

Responder is given a card:

- Scratches a random bubble in each row.
- If opposing answers are revealed, then the responder scratches out the remaining bubble in the unfavoured row.
- If same answers are revealed, then the respondent picks a row at random and scratches out the remaining bubble in that row.

Moran-Naor Physical CRRT Protocol-1

Requires scratch-off cards.

Interviewer prepares card:

- A card has four scratchable bubbles arranged in two rows of two bubbles each.
- In each row, the bubbles hide the words ‘yes’ and ‘no’.

Responder is given a card:

- Scratches a random bubble in each row.
- If opposing answers are revealed, then the responder scratches out the remaining bubble in the unfavoured row.
- If same answers are revealed, then the respondent picks a row at random and scratches out the remaining bubble in that row.

Interviewer gets back the scratched-off card:

- One row of the card is eliminated.
- Counts the result shown in the other row.

Moran-Naor Physical CRRT Protocol-1: Security Intuition

- In exactly 3/4 of the cases, the counted result will match the responder's intended result.
- Without invalidating the entire card, the responder cannot succeed with higher probability.

Moran-Naor Physical CRRT Protocol-1: Security Intuition

- In exactly 3/4 of the cases, the counted result will match the responder's intended result.
- Without invalidating the entire card, the responder cannot succeed with higher probability.

Pluasible Deniability: Responder can always claim that both rows were 'bad' and the result did not reflect the actual intention.

- Since the interviewer does not know which were the two bubbles that were scratched first, the responder's deniability cannot be refuted.

Moran-Naor Physical CRRT Protocol-1: Security Intuition

- In exactly 3/4 of the cases, the counted result will match the responder's intended result.
- Without invalidating the entire card, the responder cannot succeed with higher probability.

Pluasible Deniability: Responder can always claim that both rows were 'bad' and the result did not reflect the actual intention.

- Since the interviewer does not know which were the two bubbles that were scratched first, the responder's deniability cannot be refuted.
- Deniability is preserved even if the interviewer cheats. The only way to cheat without being unavoidable caught is to place the same answer under both bubbles in one of the rows. Considering the distribution of responses in all four cases (cheating/honest, yes/no) shows that this does not help the interviewer.

Moran-Naor Physical CRRT Protocol-1: Security Intuition

Bounded Bias: The responder learns the result before the interviewer and can decide to quit.

- Since the interviewer does not know the respondent's outcome, this has the effect of biasing the result of the poll.
- By counting the number of prematurely halted protocol executions, the interviewer can accurately estimate the number of cheating responders.

Moran-Naor Physical CRRT Protocol-1: Security Intuition

Bounded Bias: The responder learns the result before the interviewer and can decide to quit.

- Since the interviewer does not know the respondent's outcome, this has the effect of biasing the result of the poll.
- By counting the number of prematurely halted protocol executions, the interviewer can accurately estimate the number of cheating responders.

The formal proof is a detailed argument based on the above intuition.

Moran-Naor Physical CRRT Protocol-2

Requires ‘tamper-evident distinguishable’ envelopes.

Respondent prepares envelopes:

- Takes three envelopes labelled ‘1’, ‘2’ and ‘3’;
- Chooses one envelope at random and places the unfavoured answer in it;
- Places the favoured answer in the other two envelopes.

Moran-Naor Physical CRRT Protocol-2

Interviewer opens envelopes:

- Chooses a random envelope and opens it and reveals the card to the responder.
- Asks the responder to declare which of the other two envelopes contain a card with the opposite answer.

Moran-Naor Physical CRRT Protocol-2

Interviewer opens envelopes:

- Chooses a random envelope and opens it and reveals the card to the responder.
- Asks the responder to declare which of the other two envelopes contain a card with the opposite answer.
- Opens the indicated envelope.
  - If this envelope indeed has the opposite answer, then the interviewer records the answer on the first card and returns the third (unopened) envelope to the responder.

Moran-Naor Physical CRRT Protocol-2

Interviewer opens envelopes:

- Chooses a random envelope and opens it and reveals the card to the responder.
- Asks the responder to declare which of the other two envelopes contain a card with the opposite answer.
- Opens the indicated envelope.
  - If this envelope indeed has the opposite answer, then the interviewer records the answer on the first card and returns the third (unopened) envelope to the responder.
  - If both opened envelopes contain the same answer the interviewer opens the third envelope as well.

Moran-Naor Physical CRRT Protocol-2

Interviewer opens envelopes:

- Chooses a random envelope and opens it and reveals the card to the responder.
- Asks the responder to declare which of the other two envelopes contain a card with the opposite answer.
- Opens the indicated envelope.
  - If this envelope indeed has the opposite answer, then the interviewer records the answer on the first card and returns the third (unopened) envelope to the responder.
  - If both opened envelopes contain the same answer the interviewer opens the third envelope as well.
  - If the third envelope contains the opposite answer, the interviewer records the answer on the first card as the response to the poll.

Moran-Naor Physical CRRT Protocol-2

Interviewer opens envelopes:

- Chooses a random envelope and opens it and reveals the card to the responder.
- Asks the responder to declare which of the other two envelopes contain a card with the opposite answer.
- Opens the indicated envelope.
  - If this envelope indeed has the opposite answer, then the interviewer records the answer on the first card and returns the third (unopened) envelope to the responder.
  - If both opened envelopes contain the same answer the interviewer opens the third envelope as well.
  - If the third envelope contains the opposite answer, the interviewer records the answer on the first card as the response to the poll.
  - If all three envelopes contain the same answer, with probability 2/3, the interviewer records the answer appearing in the envelopes and with probability 1/3 records the opposite answer.

Moran-Naor Physical CRRT Protocol-2: Security Intuition

Bounded Bias: Responder gets her wish with probability 2/3 no matter what she does.

- Envelopes correctly filled up: the interviewer chooses the non-favourable envelope with probability 1/3.
- All envelopes have the same answer: the interviewer chooses the opposite answer with probability 1/3.

Moran-Naor Physical CRRT Protocol-2: Security Intuition

Plausible Deniability:

- The interviewer can open all three envelopes. But, then the responder can see that the seal on the returned envelope has been broken and know that the interviewer was cheating.
- The interviewer can open two envelopes before telling the responder which of the two was opened first. The interviewer hopes that the responder will not require him to return the opened envelope.
- If the penalty for cheating by the interviewer is large enough, then the responder may be convinced to answer accurately.

The formal proof is a detailed argument based on the above intuition.

Summary

- Motivation for RRT and the issue of maliciously behaving parties.
- Informal idea of security of CRRT.
- Brief descriptions of the Moran-Naor protocols for CRRT using physical devices.

Summary

- Motivation for RRT and the issue of maliciously behaving parties.
- Informal idea of security of CRRT.
- Brief descriptions of the Moran-Naor protocols for CRRT using physical devices.

Future work:

- CRRT is a meeting point of statistics and cryptography.
- Work done so far is at a basic level.
- How to capture cryptographic aspects of more complicated RRT scenarios, such as polychotomous RRT?
- Improvement of efficiencies of known protocols.
- Actual implementation and deployment of any CRRT protocol.

Thank you for your attention!