Cryptographic Randomized Response Techniques

by Ambainis, Jakobsson and Lipmaa

T-79.514, Oct 22, 2003

Ella Bingham, ella@iki.fi
HUT Lab of Computer and Information Science

Problem setting

- Elections and polls: respondents give their information only if their privacy is preserved
- A large literature exists for elections
- Polls are different in many ways:
  - less widely accepted procedures
  - less trust between the parties
  - statistical estimates are wanted instead of exact counts

  $\Rightarrow$ a strong motivation for designing poll systems

Randomized Response Techniques (RRT)

Basic setting

"Do you belong to a stigmatizing group A?"
The respondent is given a biased coin and asked to tell the truth if the coin gives heads (this has probability p_{ct}), and lie otherwise.
The a prior probability of answering "yes" is

$$p_{yes} = p_{ct} \cdot \pi_A + (1 - p_{ct})(1 - \pi_A)$$

where π_A is the overall percentage of A in the population. An unbiased estimator is $\widehat{p_{yes}} = L/N$ where L respondents out of N answer "yes".

The overall percentage of A in the population is estimated as

$$\widehat{\pi_A} = \frac{p_{ct} - 1}{2p_{ct} - 1} + \frac{L}{N} \cdot \frac{1}{2pct - 1}.$$

We will say that a respondent is of type $t = 1$ if she belongs to group A, and $t = 0$ otherwise.

Innocuous question method

The respondent is given two questions: the one of interest in the poll, and another completely harmless. She chooses between the two questions by a toss of a biased coin.

Polychtomous RRT

A question with multiple mutually exclusive answers $A_1, \ldots, A_m$, some of which are harmless and some of which the respondent typically wants to keep as a secret.

Problems with RRT

The respondent may not want to lie, even if asked to. Or she may refuse to answer to some questions. This biases the estimation of π_A. To overcome this, the authors propose Cryptographic RRT.

Cryptographic RRT

Guarantees the privacy of the respondent

Also guarantees the privacy of the interviewer: the respondent cannot determine the outcome of the protocol before the end.

Some basics of algebra

A *group* is a set G together with some operation $*$ which obeys

- If $a, b \in G$ then $a * b \in G$
- $(a * b) * c = a * (b * c)$
- There is an identity element I such that $I * a = a * I = a \, \forall a \in G$
- Every $a \in G$ has an inverse a^{-1} such that $a * a^{-1} = a^{-1} * a = I$

We will use $\mathbb{Z}_p$ which is the set of integers modulo an integer p: $\mathbb{Z}_p = \{0, \ldots, p-1\}$. In other words, if we

divide any integer by p then the remainder is in $\mathbb{Z}_p$.

Let G be a group and $g \in G$. Let $\langle g \rangle = \{g^i | i \geq 0\}$ be the set of the powers of g. We say that g is a *generator* of G if $\langle g \rangle = G$.

For example, consider $G = \{1, 2, 4, 5, 7, 8\} \subset \mathbb{Z}_9$. 2 is a generator of G:
$\langle 2 \rangle = \{2^0, 2^1, 2^2, 2^3, 2^4, 2^5, \ldots\} = \{1, 2, 4, 8, 7, 5\}$

If g is a generator, then for any $y \in G$ there is a unique $i \in \{0, \ldots, m-1\}$ (where m is the number of elements in G) such that $g^i = y$. This i equals $\log_g(y)$ and takes exponential time to find.

Protocol 1

Background

- p and q are primes such that q divides $p - 1$. The public key consists of g and h that are two generators of G that is a unique subgroup of $\mathbb{Z}_p$.
- Even if g and h are known, $g^{\mu}h^{v}$ is hard to invert (here μ is the message, and v is picked at random from $\mathbb{Z}_q$).
- $n, \ell \in \mathbb{N}$ such that $p_{ct} = \ell/n > 1/2$.

Precomputation step:

- The respondent $\mathcal{R}$ prepares n random bits $\mu_i \in \{0, 1\}$ for $i = 1, \ldots, n$, such that $\sum_i \mu_i = \ell$ if her type is $t = 1$ and $\sum_i \mu_i = n - \ell$ if $t = 0$.
  (Thus $p_c t = \ell/n$ is the probability that a randomly picked bit equals her type).
  Additionally, she sets $\mu_{n+1} \leftarrow t - 1$.
- The interviewer $\mathcal{I}$ chooses $\sigma \in \{1, \ldots, n\}$

Interactive step:

- $\mathcal{I}$ picks a and b at random from $\mathbb{Z}_q$ and sends g^a, g^b and $g^{ab-\sigma+1}$ to $\mathcal{R}$.
- $\mathcal{R}$ repeats the following for all $i \in \{1, \ldots, n\}$: Pick r_i and s_i at random from $\mathbb{Z}_q$. Compute
  $w_i \leftarrow g^{r_i}(g^a)^{s_i} = g^{r_i + a s_i}$
  and
  $v_i \leftarrow (g^b)^{r_i}(g^{ab-\sigma+1}g^{i-1})^{s_i} = g^{(r_i+as_i)b+(i-\sigma)s_i}$,
  and use v_i as a key to encrypt the answer μ_i to y_i using
  $y_i \leftarrow g^{\mu_i} h^{v_i}$.
  Send w_i and y_i to $\mathcal{I}$.
- $\mathcal{I}$ computes w_σ^b (note that when $i = \sigma$ above, then

the key v_i is w_i^b) and
$g^{\mu\sigma} \leftarrow y_\sigma / h^{w_\sigma^b}$
and then computes μ_σ from that.
(With probability p_{ct}, this is 1, and he will conclude $r_{\mathcal{R}} = 1$; with probability $1 - p_{ct}$, this is 0 and $r_{\mathcal{R}} = 0$.)

- $\mathcal{R}$ must now prove that she created $\{\mu_1, \ldots, \mu_{n+1}\}$ correctly. Use noninteractive zero-knowledge arguments (details are seen in the paper).
- $\mathcal{I}$ verifies the arguments, and halts if the verification fails.

The interviewer's output r_{calR} corresponds to the "yes" answer in the basic RRT: in computing π_A, L is now the number of $r_{\mathcal{R}} = 1$ values in the population.

Protocol 2

Now $d = \lceil 1/(1 - p_{ct}) \rceil$, other background is as before.

Precomputation:

- $\mathcal{R}$ chooses a random $\mu \in \{0, 1, \ldots, n - 1\}$.
- $\mathcal{I}$ chooses random $\nu \in \{0, 1, \ldots, n - 1\}$ and $\sigma \in \{0, 1, \ldots, d - 1\}$.

Interactive step:

- $\mathcal{R}$ commits to t and μ and sends the commitments to $\mathcal{I}$.
- $\mathcal{I}$ chooses a random ρ and commits to σ by setting $y \leftarrow C_K(\sigma; \rho)$. He sends ν and y to $\mathcal{R}$, together with a zero-knowledge argument for y.
- $\mathcal{R}$ verifies the argument. She computes for all $i \in \{0, 1, \ldots, d-1\}$ a value μ'_i such that $\mu'_i = t$ if and only if $(\mu + \nu + i\ell \mod n) < \ell$. She signs y and sends her signature together with all μ_i and a zero-knowledge argument.
- $\mathcal{I}$ sets $r_{\mathcal{R}} \leftarrow \mu'_\sigma$, accompanied with $\mathcal{R}$'s signature on the commitment.

Quantum cryptographic RRT

- allows using p_{ct} that is not a rational number
- provides a relaxed form of information-theoretic security for both parties:
  - if $\mathcal{R}$ is dishonest, her vote only counts as $\leq \sqrt{2}$ votes
  - if $\mathcal{I}$ gets to know $\mathcal{R}$'s private input with some probability, he is also caught cheating with another probability.
- the protocol can implemented using contemporary technology