We want to give access to polls and survey to the regular developer.

This is a LINQ query for asking college students whether they are liberal arts majors.

We want to make access to human-generated data as easy as access to databases.
```csharp
var femaleHeight = from person in people where person.Gender == Gender.FEMALE select person.PoseQuestion<int>("What is your height?");

var maleHeight = from person in people where person.Gender == Gender.MALE select person.PoseQuestion<int>("What is your height?");

if (maleHeight.ToRandomVariable() > femaleHeight.ToRandomVariable()) {
    Console.WriteLine("Males are taller that females, according to a t-test.");
}
```
LINQ (Language-integrated queries) is the only way to create surveys, but it has several important advantages.

In InterPoll, we are thinking a lot about the developer experience. LINQ is familiar to developers and does not require learning a new domain-specific language.

LINQ queries integrate well with the rest of the application.
using ...

namespace Microsoft.Research.RiSE.InterPoll
{
    public partial class Runner
    {
        [TestMethod]
        public void EmploymentSurvey()
        {
            // Implementation goes here
        }
    }
}
Sample Surveys

Survey

Do you shop locally?
- Always
- Never
- Once in a while
- Usually
- About half the time

Do you make at least one purchase a day at chain stores?
- Yes
- No

Do you shop at local stores daily?
- Yes
- No

Do you consider yourself to be a supporter of small business?
- Yes
- No

Survey

1. I feel tense or "wound up"
- Most of the time
- A lot of the time
- From time to time, occasionally
- Not at all

2. I still enjoy the things I used to enjoy
- Definitely as much
- Not quiet so much
- Only a little
- Hardly at all

3. I get a sort of frightened feeling as if something awful is about to happen
- Very definitely and quiet badly
- Yes, but not too badly
- A little, but it doesn't worry me
- Not at all

4. I can laugh and see the funny side of things
- As much as I always could
- Not quite so much now
- Definitely not so much now
- Not at all

5. Worrying thoughts go through my mind
- A great deal of the time
- A lot of the time
- From time to time but not too often
- Only occasionally

6. I feel cheerful
- Not at all
- Not often
- Sometimes
InterPoll uses Mechanical Turk as a back-end

Others like UHRS are possible
IS THIS NOT A SOLVED PROBLEM?

SurveyMonkey

Create surveys anywhere, anytime.

Analyze your survey results on the go.

Track your results in real-time.

Customer Feedback Survey

1. How well did the customer service agent answer your questions?
   - Extremely well
   - Very well
   - Moderately well
   - Slightly well
   - Not at all well

Add Question
Add Page

SHOWING: 312 of 312 Responses

1. How well did the customer service agent answer your questions?

- Extremely well: 18% (56 responses)
- Very well: 35% (110 responses)

312 Total Responses
288 Completed Responses
Status: OPEN

Edit
Send
Analyze

Completion Rate: 75%
OUR FOCUS IS ON AN END-TO-END PROCESS

1) idea
2) poll
3) crowd
4) analysis
KEY DIFFERENCES BETWEEN THEM AND INTERPOLL

- Lack of programmability
- InterPoll is considerably cheaper
- We give statistical significance to the results through unbiasinging and power analysis
A SMORGASBORD OF TOPICS

- Crowd-sourcing
- Human and machine computation combined
- Programming with uncertainty
- Privacy
- Statistics
- Database and PL optimizations
One of the wonderful things about InterPoll is that it brings together research fields that are not frequently combined.

This requires researchers who are not afraid to step outside their comfort zone.
“Experimental psychology is the study of the college sophomore”

Quinn McNemar, 1946
Traditional polling is a painstaking and time-consuming process.

Not everything can be assessed through computer-based polling, but many things can be.

Recent studies show that online polling is frequently better than RDD polls and are good at avoiding the interviewer bias.
Recent results suggest that online polling is actually cheaper, faster, and is devoid of some of the biases such as the interviewer bias present in traditional face-to-face or telephone surveys.

Moreover, in recent years, traditional random-dialing surveys suffer from the fact that young people often do not have a land line.
COSTLY... ESPECIALLY AT SCALE

SurveyMonkey Audience pricing

A technology accessory company wants feedback on their latest iPhone case design. They have a 14-question survey and would like 200 responses from an audience of only female iPhone users, delivered on the 2-business day schedule.

<table>
<thead>
<tr>
<th>Survey Feature</th>
<th>Cost per Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 question survey</td>
<td>$1.50 per response</td>
</tr>
<tr>
<td>2 specific targeting options added</td>
<td>$1.25 per response (Gender targeting &amp; iPhone ownership targeting)</td>
</tr>
<tr>
<td>2-business day turnaround</td>
<td>$1.00 per response</td>
</tr>
<tr>
<td>Project cost for 200 responses</td>
<td>$750.00 ($3.75 per response)</td>
</tr>
</tbody>
</table>

Instant.ly cost per completed survey

![Cost per completion, in $ vs. Number of questions asked](chart.png)

Cost per completion, in $ vs. Number of questions asked

Number of questions asked

Cost per completion, in $
Our main backend is Mechanical Turk

Generally can get answers to most polls for $.10/survey completion

Latency drops as a result of higher rewards, but the quality generally doesn’t seem to increase

InterPoll takes care of creating appropriate HITs, generating XML forms for the UI that appears on Mechanical Turk and collecting the results
How to integrate human and computer computation is a wide open problem in programming languages.

Our approach is to use LINQ to provide seamless language extensions.

But many topics remain unaddressed. How does one program with long-running crowd operations?

How does one combine lazy and eager computation?
Uncertain<T>
24 mph
59 mph
A WALK WITH A GPS
HOW DO YOU WRITE THIS KIND OF CODE?

GeoCoordinate Location;
HOW DO YOU WRITE THIS KIND OF CODE?

```cpp
GeoCoordinate Location;

Uncertain<GeoCoordinate> Location;
```
Uncertain\texttt{<T>} is an uncertain type abstraction.

It encourages developers to explicitly reason about uncertainty.
A variable of type Uncertain<T> is a random variable, represented by a distribution.
Uncertain<\texttt{double}> Speed = Dist / 5;

Or more generally, $Z = X + Y$, if $X$ and $Y$ are distributions.

\begin{itemize}
  \item If $x$ is a sample of $X$
  \item and $y$ is a sample of $Y$
  \item then $x+y$ is a sample of $X+Y$ *
\end{itemize}

* if $X$ and $Y$ are independent
Sampling function for E recursively samples children.
if (Speed > 4) print("Great job!");
if (Speed > 4) print("Great job!");
if (Speed > 4) print("Great job!");

More likely than not that Speed > 4?

Pr[Speed > 4] > 0.5?
TESTING DISTRIBUTIONS

```python
if (Speed > 4) print("Great job!");
```

At least 90% likely that Speed > 4?
testing distributions

if (Speed > 4).Pr(0.9) print("Great job!");

null hypothesis $H_0$: \( \Pr[\text{Speed} > 4] \leq 0.9 \)

alternate hypothesis $H_A$: \( \Pr[\text{Speed} > 4] > 0.9 \)

approximate!

How many samples?

Too many = too slow

Too few = too noisy

Sequential sampling: sample size depends on progress
A and B depend on X — not independent — take one X sample not two!
Uncertain language semantics implements Bayes Rule

\[ \Pr[H|E] = \frac{\Pr[E|H] \Pr[H]}{\Pr[E]} \]
GPS SPEED
GPS SPEED
<table>
<thead>
<tr>
<th>Time</th>
<th>GPS speed (95% CI)</th>
<th>Improved speed (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**GPS SPEED**
Uncertain<T> is an uncertain type abstraction. It encourages developers to explicitly reason about uncertainty.
BRINGING IT ALL TOGETHER

sensors

people
SOCIAL SCIENCES

What is the prevalence of depression in men vs. women?

How near-sighted are people over 50?

Which video game console do you prefer? Xbox One/PS4/Both/Neither?

What do you think about doctor-assisted suicide?

Do you think affirmative action as part of college admission is a good idea?

Do you itemize your taxes or use the standard tax form?

What do you think about race relations in America today?
HEALTHCARE SURVEYS

- How many hours of sleep have you gotten on average in the past month?
- Do you feel that you get enough sleep?
- How is duration of sleep correlated with coffee intake for college students?
- Are parents satisfied with the amount of sleep they get?
- Are men more satisfied than women?
- For people between 30-45, do they find medical providers by word of mouth?
- Which is worse for your health: alcohol, tobacco, sugar, or marijuana?
- Is Yelp a trusted source of healthcare recommendations?
Are men or women responsible for shopping for pet food?

Do people between 20-30 play more video games than those between 30-45?

Which video game console do you prefer?

What kind of jeans do you wear? (Skinny/straight/loose, etc.)

What percentage of the population between 30 and 50 use both cable and streaming services like Netflix?

Do you like the new Microsoft company logo better than the old Microsoft company logo?

Have you paid for a cable TV subscription in the last year and, if so, are you likely to have a Hulu subscription?
POLITICAL POLLING

In cases of extreme drought, should farmers and food producers have priority when allocating water?

Should individuals obtain permission to photograph every person in the photos they take?

Are you in favor or against raising the federal minimum wage?

Do you think Obamacare will make things better or worse for you and your family?

Should legalized recreational marijuana be required to be sold with a warning label?

How did you or will you pay for higher education (anything post-high school)?

How do you feel about the direction of the country?
Demo: basic surveys
OTHER INTERFACES
FUNDAMENTAL PROBLEMS
MOVING AWAY FROM SMALL AND UNREPRESENTATIVE SAMPLES
POWER ANALYSIS

Determine the number of samples for a query

We can sample from the crowd sequentially until we satisfy or disprove our hypothesis.

We will poll the crowd for more until our stopping criterion is reached.

The stopping criterion allows us to conclude that the hypothesis can be proven or disproven with the required level of confidence.
EXAMPLE QUESTION: HEIGHT

N=29

Once we remove the outliers

height = from person in height where
where

  person.Height >=140 &&

N=27
ARE THE TWO CORRELATED?

Hospital anxiety and depression scale (HADS)
- Normal 0-7
- Mild 8-10
- Moderate (11–14)
- Severe (15–21)

```
var happinessData = from person in people
select new
{
    Consider = person.PoseCodedQuestion("I feel tense or 'wound up'", new Tuple<string, string>());
    Enjoy = person.PoseCodedQuestion("I still enjoy the things I used to enjoy", new Tuple<string, string>());
    Awful = person.PoseCodedQuestion("I get a sort of frightened feeling as if something will happen", new Tuple<string, string>());
    Laugh = person.PoseCodedQuestion("I can laugh and see the funny side of things", new Tuple<string, string>());
    Worry = person.PoseCodedQuestion("Worrying thoughts go through my mind", new Tuple<string, string>());
    Cheerful = person.PoseCodedQuestion("I feel cheerful", new Tuple<string, string, int>()("Not at all", "Slight", "A lot"));
    Sit = person.PoseCodedQuestion("I can sit at ease and feel relaxed", new Tuple<string, string>());
    Slowed = person.PoseCodedQuestion("I feel as if I am slowed down", new Tuple<string, string>());
    Frightened = person.PoseCodedQuestion("I get a sort of frightened feeling like 'butterflies' in my stomach", new Tuple<string, string>());
};
```

```
var scores = from data in happinessData
select new
{
    Anxiety = data.Consider + data.Awful + data.Worry + data.Sit + data.Frightened + data.Restless + data.Panic,
    Depression = data.Enjoy + data.Laugh + data.Cheerful + data.Slowed + data.LostInterest + data.LookForward + data.GoodBook,
    Gender = data.Gender,
    Income = data.Income,
    Education = data.Education,
    Ethnicity = data.Ethnicity,
    Employment = data.Employment,
};
```
DOES MONEY BUY HAPPINESS (OR AT LEAST TRANQUILITY)?

Are rich more anxious than poor?

\[ N = 105 \]

expected value for poor = 8.5714, expected value for rich = 7.9619
# Millennials Don't Stand A Chance

## DEBATE DETAILS

### THE PANEL

### RESULTS

<table>
<thead>
<tr>
<th>Task</th>
<th>Outcome</th>
<th>Power</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Millennials Don't Stand A Chance</td>
<td>No</td>
<td>37</td>
<td>$3.70</td>
</tr>
<tr>
<td>Minimum Wage</td>
<td>No</td>
<td>43</td>
<td>$4.30</td>
</tr>
<tr>
<td>Rich Are Taxed Enough</td>
<td>No</td>
<td>51</td>
<td>$5.10</td>
</tr>
<tr>
<td>End Of Life</td>
<td>No</td>
<td>53</td>
<td>$5.30</td>
</tr>
<tr>
<td>Break Up The Big Banks</td>
<td>Yes</td>
<td>73</td>
<td>$7.30</td>
</tr>
<tr>
<td>Strong Dollar</td>
<td>No</td>
<td>85</td>
<td>$8.50</td>
</tr>
<tr>
<td>Marginal Power</td>
<td>No</td>
<td>89</td>
<td>$8.90</td>
</tr>
<tr>
<td>Genetically Engineered Babies</td>
<td>Yes</td>
<td>135</td>
<td>$13.50</td>
</tr>
<tr>
<td>Affirmative Action On Campus</td>
<td>Yes</td>
<td>243</td>
<td>$24.30</td>
</tr>
<tr>
<td>Obesity Is Government Business</td>
<td>No</td>
<td>265</td>
<td>$26.50</td>
</tr>
</tbody>
</table>

Illustration by Thomas James

Wednesday, April 9, 2014
CONVERGENCE CURVES: SEQUENTIAL PROBABILITY RATIO TEST (SPRT) OR WALD, 1945

Sequential probability ratio test: To implement this, we build a sequential acceptance plan. Let $H_0$: $p + \epsilon$ and $H_A: p - \epsilon$ where $p = 0.5$ by default and can be overloaded by a programmer. Uncertain(T) calculates the cumulative log-likelihood ratio for each sample:

$$\Delta_L = k \log(H_A/H_0) + (n - k) \log(H_0/H_A)$$

where $n$ is the number of samples taken thus far and $k$ is the number of successes out of those $n$ trials. If

$$\Delta_L \leq \log(\alpha/(1 - \alpha)) = a$$

then Uncertain(T) evaluates the conditional as false while if

$$\Delta_L \geq \log((1 - \alpha)/\alpha) = b$$

the conditional is true.
BE CAREFUL WITH YOUR PARAMETERS

As we vary the probability

As we vary the confidence
TWO MORE CONUNDRUMS

Affirmative action does more harm than good

We should break up the big banks
PRIORS FOR THE CROWD

Instant.ly crowd

36% Male
64% Female

US census

49% Male
51% Female
SEGMENTS OF THE CROWD LOOK DIFFERENT

Panos Ipeirotis, 2010
THE UNBIAS OPERATOR

```csharp
var photoAttitudes = (from person in people
                      select new
                      {
                        Used = person.PoseQuestion<bool>(
                        "Have you ever hired a professional photographer?"),
                        WorthIt = person.PoseQuestion<bool>(
                        "Do you feel the money you spent was worth the experience?"),
                        Quality = person.PoseQuestion(
                        "How would you rate the quality of the pictures?"
                        "★", "★★", "★★★")
                        HowLikely = person.PoseQuestion(
                        "How likely are you to recommend this service to others?"
                        "Very likely", "Somewhat likely", "Not likely at all")
                        WhatDidYouEnjoy = person.PoseQuestion("What did you most enjoy about the experience?")
                      });

  // priors for demographics.
  var mturk = MTurkPriors.DefaultPriors;
  var census = CSPSlicedPriors.DefaultPriors;

  var correctedAttitudes = Unbiasing.Unbias(photoAttitudes,
                                           p => p.HowLikely,
                                           p => p.Gender, mturk, census);
```
UNBIASING IN INTERPOLL

- I'm not sure: Biased response 0.100, Unbiased response 0.100
- Somewhat likely: Biased response 0.350, Unbiased response 0.350
- Somewhat unlikely: Biased response 0.100, Unbiased response 0.100
- Very likely: Biased response 0.300, Unbiased response 0.300
- Very unlikely: Biased response 0.050, Unbiased response 0.050
LOCAL OPTIMIZATIONS

Poll twice?
Poll once and reuse the population?

Dead code – reason about collection properties and query/code interactions
**CONDITIONAL BLENDING**

```csharp
var unemployed =
    from person in people
    where person.Employment==Employment.WORKED_FULL_TIME
    select person;

foreach (var person in unemployed)
{
    if (person.Gender == Gender.FEMALE)
    {
        Console.WriteLine("unemployed female: {0}", person);
    }
}

var unemployedFemales =
    from person in people
    where
        person.Employment==Employment.WORKED_FULL_TIME_YEAR
        person.Gender == Gender.FEMALE
    select person;

foreach (var person in unemployedFemales)
{
    Console.WriteLine("unemployed female: {0}", person);
}
```
CROSS-CROWD SAMPLING STRATEGIES

Having more than one crowd can help us assess the issue of ignorability: does the fact that we are polling online affect the outcome?

Test respective biases brought about by different crowds

Provide answers for different population mixes: 70% Indian, 30% American

Specific optimizations

- US -> India
- US and India, blended
- US during the day, India at night
- Use India to find missing population segments
MORE OPTIMIZATIONS

Can we cache answers to common questions and reuse them across surveys? Is that always the correct thing to do?

Can we determine questions that are time-sensitive?

Can we decide if keeping answers to specific questions may lead to privacy violations.

Can we determine questions suitable for different crowds?

Can we determine if someone is likely to cheat or compensate for the possibility of cheating?
EXPLORING THE DATA
using ...

namespace Microsoft.Research.RiSE.InterPoll
{
    public partial class Runner
    {
        [TestMethod]
        public void EmploymentSurvey()
        {
        }
    }
}
CONCLUSIONS

InterPoll: a system for large-scale crowd-sourced polling

Geared toward
  • Developers who want to incorporate human data into their applications
  • But also social scientists
  • Marketing professionals
  • Campaign pollsters

Have explored power analysis and are doing experiments on unbiasing and various optimizations