Uncertain<$T>$
A First-Order Type for Uncertain Data

James Bornholt  Australian National University
Todd Mytkowicz  Microsoft Research
Kathryn S. McKinley  Microsoft Research
Location Loc = GetGPSLocation();

estimated data

discrete type
Uncertainty bug applications use estimated data, but languages use discrete types

estimated data

discrete type

Location Loc = GetGPSLocation();
public class GeoCoordinate {
  public double Latitude;
  public double Longitude;
  
  public double HorizontalAccuracy;
}
Using estimates as facts

```java
public class GeoCoordinate {
    public double Latitude;
    public double Longitude;
    public double HorizontalAccuracy;
}
```
Using estimates as facts

```java
public class GeoCoordinate {
    public double Latitude;
    public double Longitude;
    public double HorizontalAccuracy;
}

OS #1
95% confidence interval
σ = 33 m

OS #2
68% confidence interval
σ = 39 m
Using estimates as facts

Computation compounds error

Uncertain<\text{T}>: A First-Order Type for Uncertain Data
Using estimates as facts

Computation compounds error

Inference asks wrong questions

\textbf{if (Distance < 200) ...}
How do application writers handle estimated data without a PhD in statistics?
How do application writers handle estimated data without a PhD in statistics?

**Uncertain<T>** is an uncertain type abstraction.
Related work

Developer computations

Probabilistic data
Sensors, measurements, probabilistic models
Related work

Developer computations

No abstraction

Domain PhD

Probabilistic data
Sensors, measurements, probabilistic models
Related work

- No abstraction
  - Domain PhD
- Probabilistic data
  - Sensors, measurements, probabilistic models

Developer computations

- Probabilistic programming
- Statistics PhD
Related work

- No abstraction
  - Domain PhD
  - Probabilistic data
    - Sensors, measurements, probabilistic models

- Probabilistic programming
  - Statistics PhD

- Current abstractions
  - App developers

Developer computations
## Related work

<table>
<thead>
<tr>
<th>Developer computations</th>
</tr>
</thead>
<tbody>
<tr>
<td>No abstraction</td>
</tr>
<tr>
<td>Probabilistic programming</td>
</tr>
<tr>
<td>Uncertain&lt;T&gt;</td>
</tr>
<tr>
<td>Current abstractions</td>
</tr>
</tbody>
</table>

- **Domain PhD**
- **Statistics PhD**
- **App developers**

**Probabilistic data**
- Sensors, measurements, probabilistic models
Related work

Developer computations

No abstraction

Flexible

Probabilistic programming

Statistics PhD

Probabilistic data

Sensors, measurements, probabilistic models

Current abstractions

Simple

App developers

App developers

Uncertain<T>

Uncertain<T>: A First-Order Type for Uncertain Data
Identifying distributions

Experts provide measurement models (they probably already have them!)

Uncertain\(<T>\) approximates distributions by random sampling
Uncertain\( <T> \) lifts arithmetic operators (like +) to work over distributions.

Sampling trades speed for accuracy.
“Is distance less than 200 m?”

The answer is a probability.

Uncertain\(<T>\) accounts for sampling error using hypothesis testing on expected values.
Bayes’ Theorem uses distributions to form better estimates

\[
Pr[H|E] = \frac{Pr[E|H] \cdot Pr[H]}{Pr[E]}
\]

Identifying distributions
Computing with estimates
Asking the right questions
Improving estimates
Uncertainty bugs: applications use estimated data, but languages use discrete types.

Uncertain<\(T\)> is an uncertain type abstraction.

Uncertain<\(T\)> makes programs more expressive and more correct.