

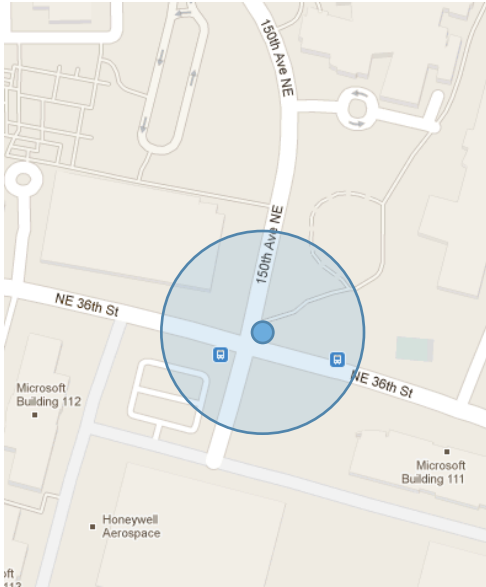
UncertainT>

A First-Order Type for Uncertain Data

James Bornholt Australian National University

Todd Mytkowicz Microsoft Research

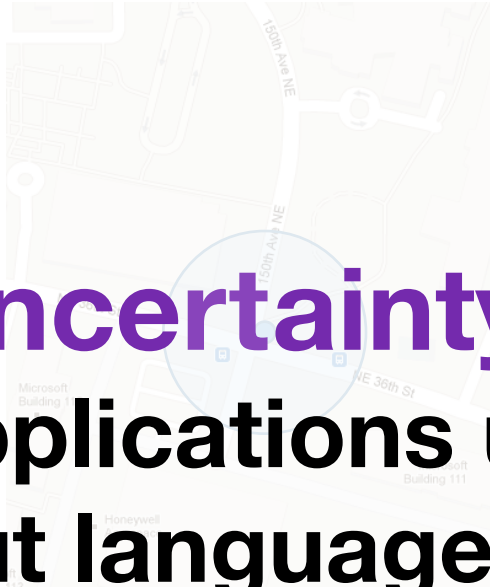
Kathryn S. McKinley Microsoft Research



**estimated
data**

Location Loc
= GetGPSLocation();

**discrete
type**



Uncertainty bug

```
Location Loc  
= GetGPSLocation();
```

**applications use estimated data,
but languages use discrete types**

estimated
data

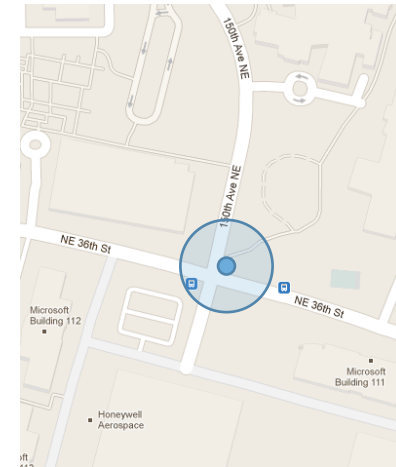
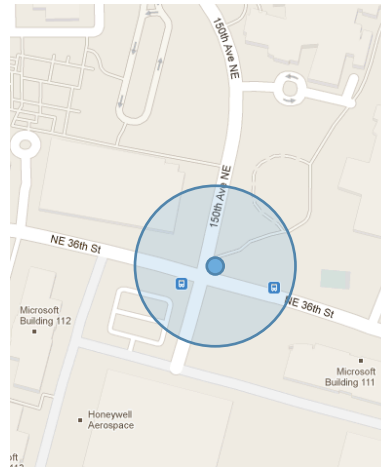
discrete
type

Using estimates as facts

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

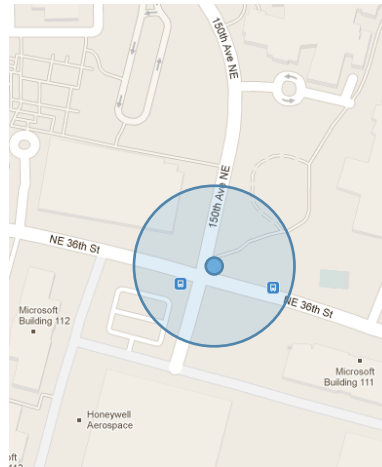
Using estimates as facts

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



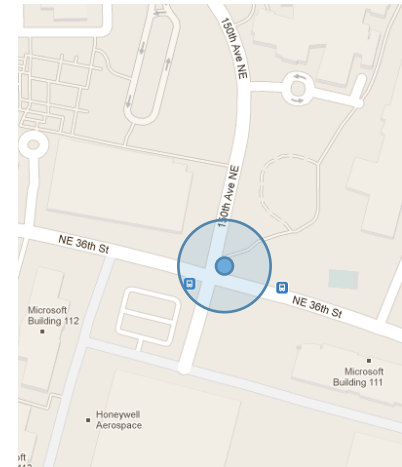
Using estimates as facts

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



OS #1

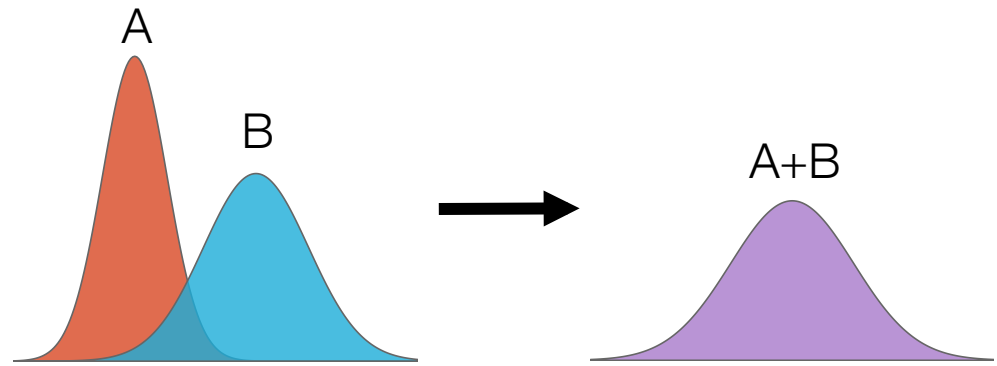
95% confidence interval
 $\sigma = 33$ m



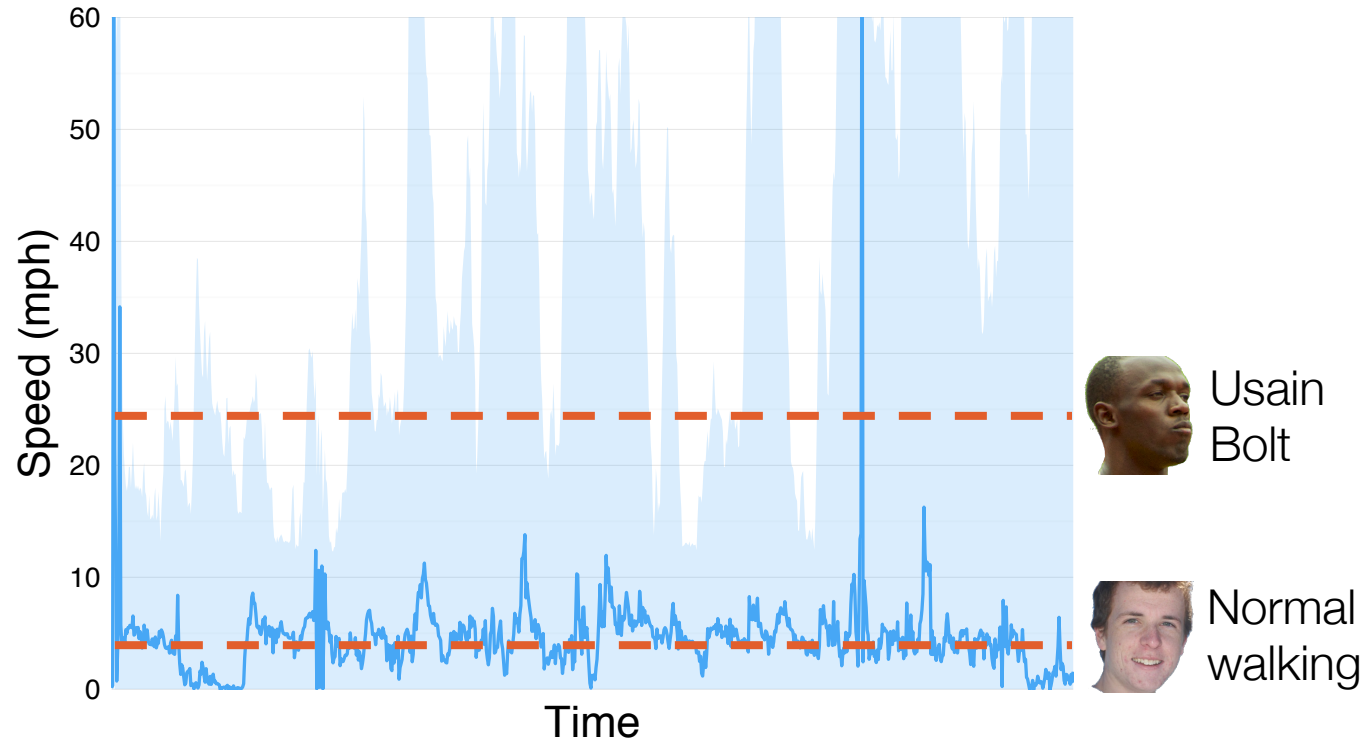
OS #2

68% confidence interval
 $\sigma = 39$ m

Using estimates as facts



Computation compounds error

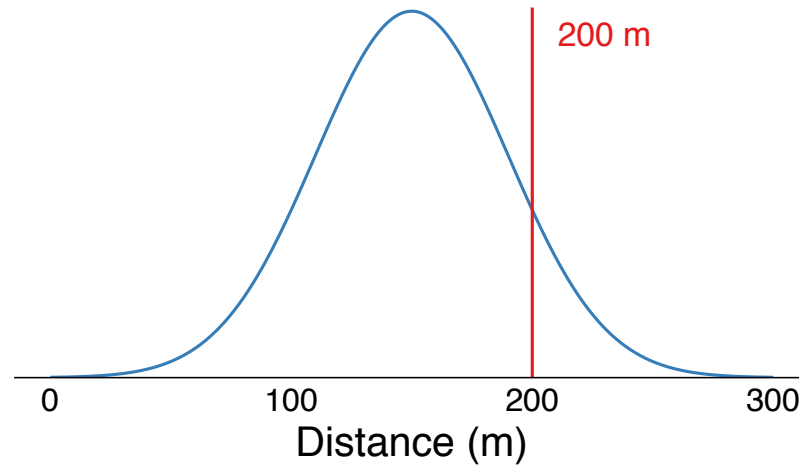


Using
estimates as
facts

```
if (Distance < 200) ...
```

Computation
compounds
error

Inference
asks wrong
questions



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

Developer computations

No abstraction

Probabilistic
programming

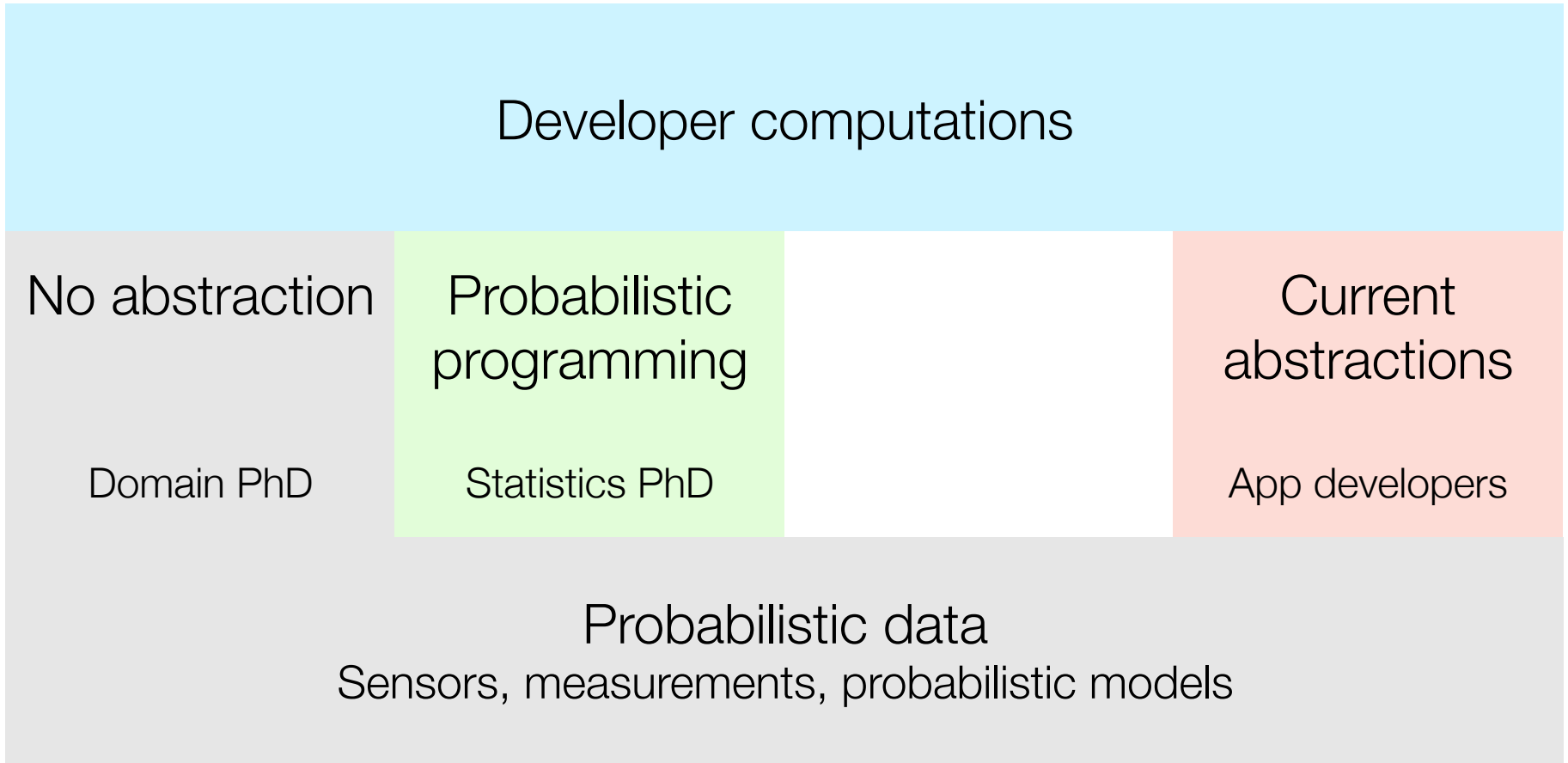
Domain PhD

Statistics PhD

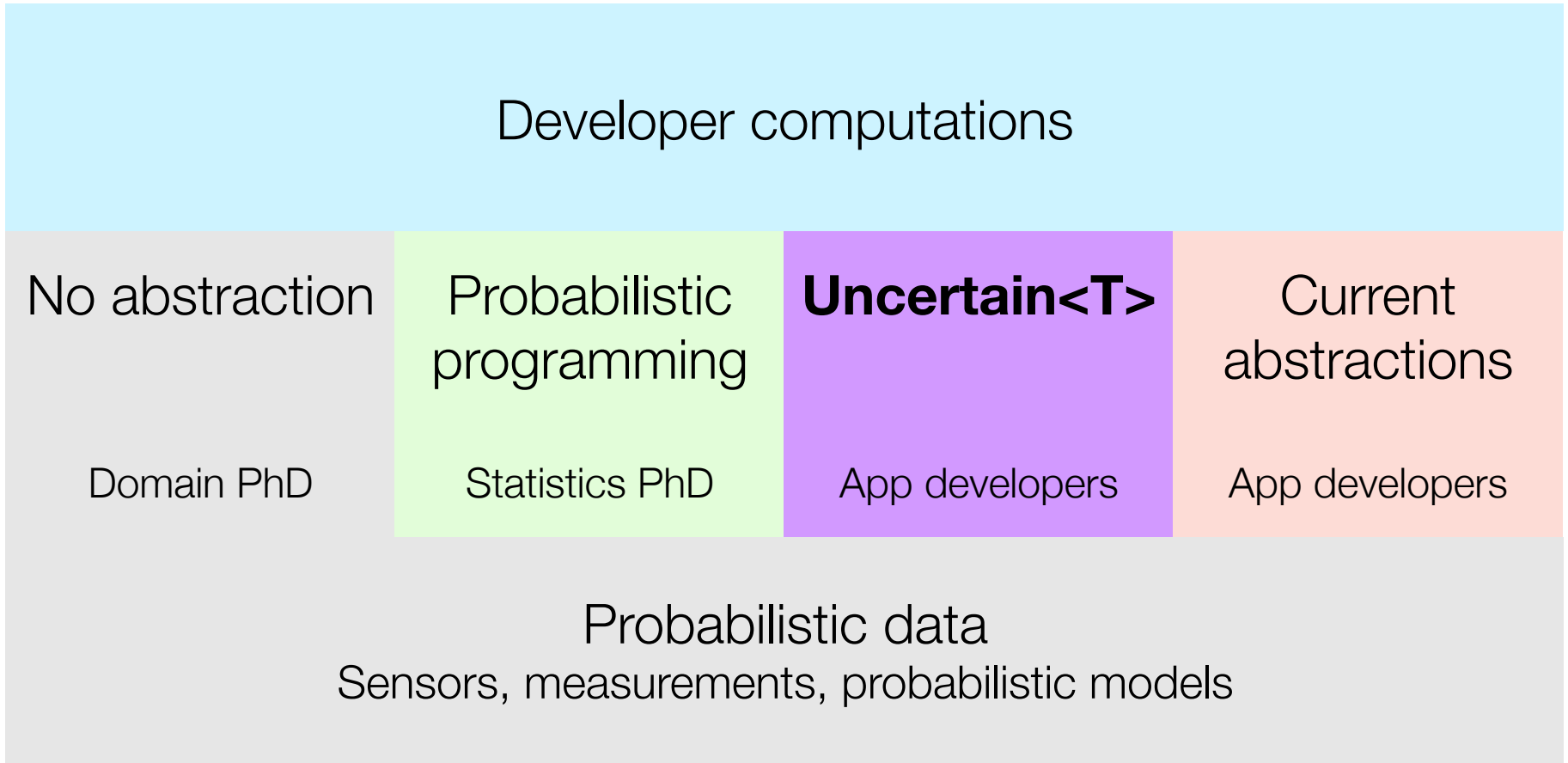
Probabilistic data

Sensors, measurements, probabilistic models

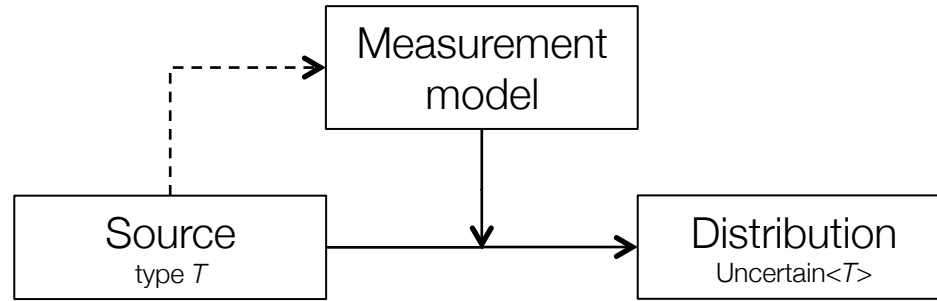
Related work



Related work



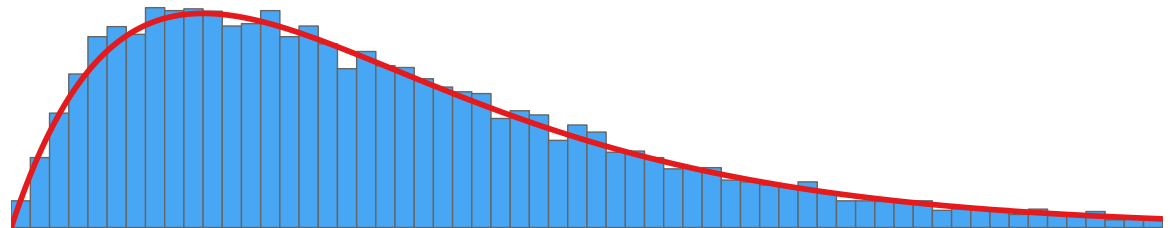
1 Identifying distributions



2 Computing with estimates

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

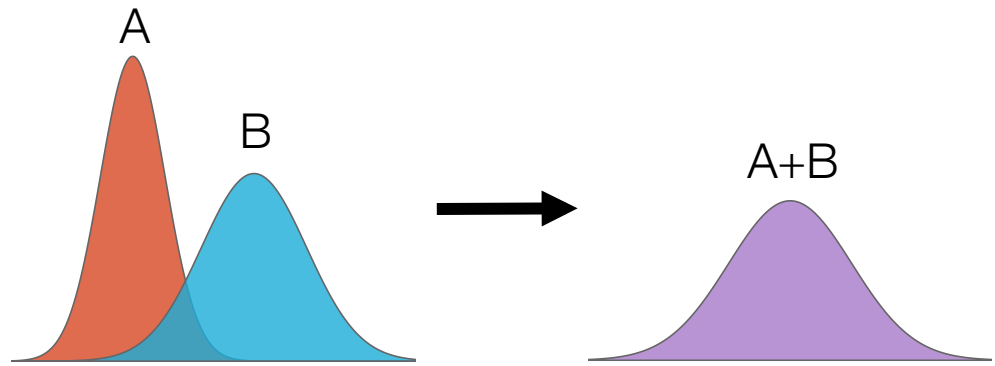
3 Asking the right questions



4 Improving estimates

Uncertain $\langle T \rangle$ approximates distributions by random sampling

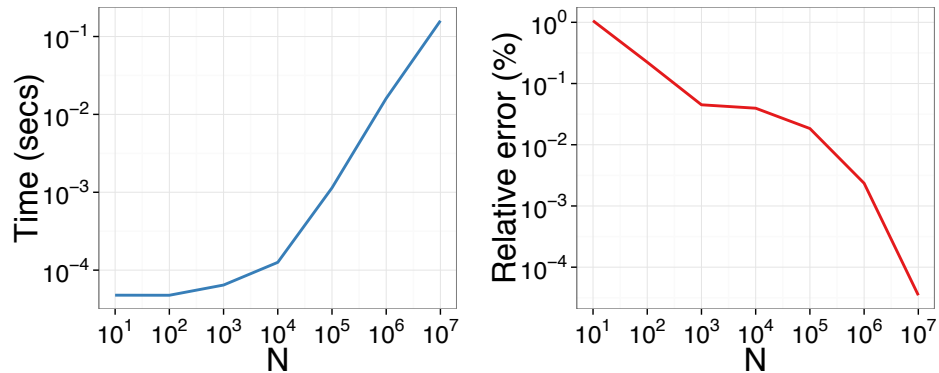
1 Identifying distributions



2 Computing with estimates

Uncertain $\langle T \rangle$ lifts arithmetic operators (like +) to work over distributions

3 Asking the right questions

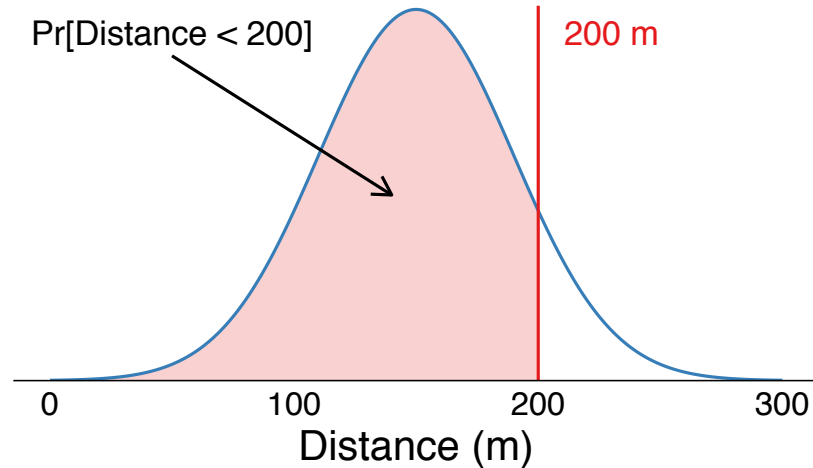


4 Improving estimates

Sampling trades speed for accuracy

1 Identifying distributions

“Is distance less than 200m?”



2 Computing with estimates

3 Asking the right questions

The answer is a probability.

4 Improving estimates

Uncertain $\langle T \rangle$ accounts for sampling error using hypothesis testing on expected values.

1 Identifying distributions

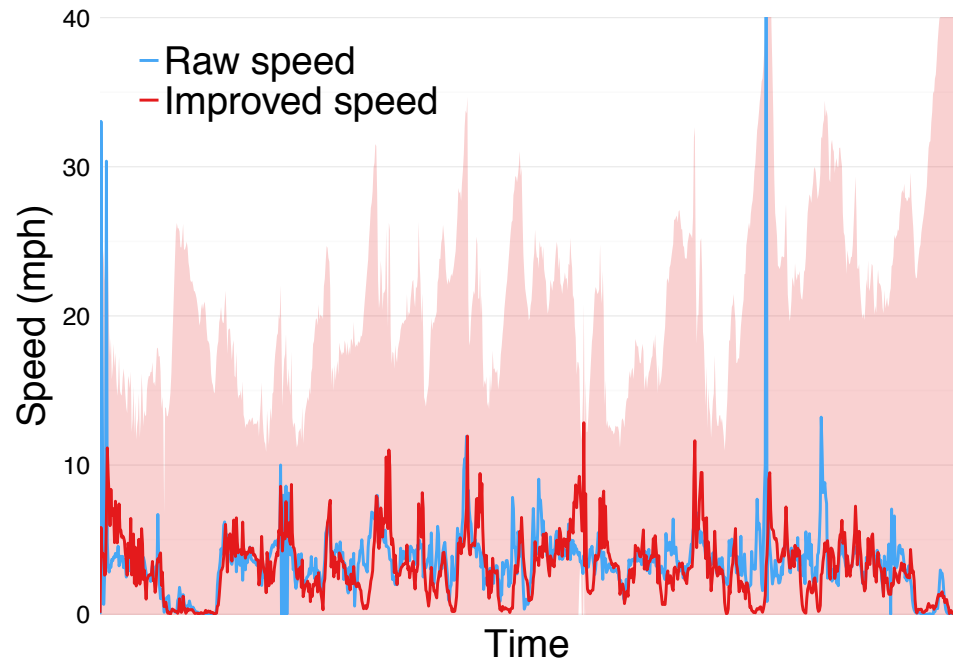
$$\Pr[H|E] = \frac{\Pr[E|H] \Pr[H]}{\Pr[E]}$$

Bayes' Theorem uses distributions to form better estimates

2 Computing with estimates

3 Asking the right questions

4 Improving estimates



Uncertainty bugs: applications use estimated data, but languages use discrete types.

Uncertain $\langle T \rangle$ is an uncertain type abstraction.

Uncertain $\langle T \rangle$ makes programs more expressive and more correct.