# Debugging Probabilistic Programs

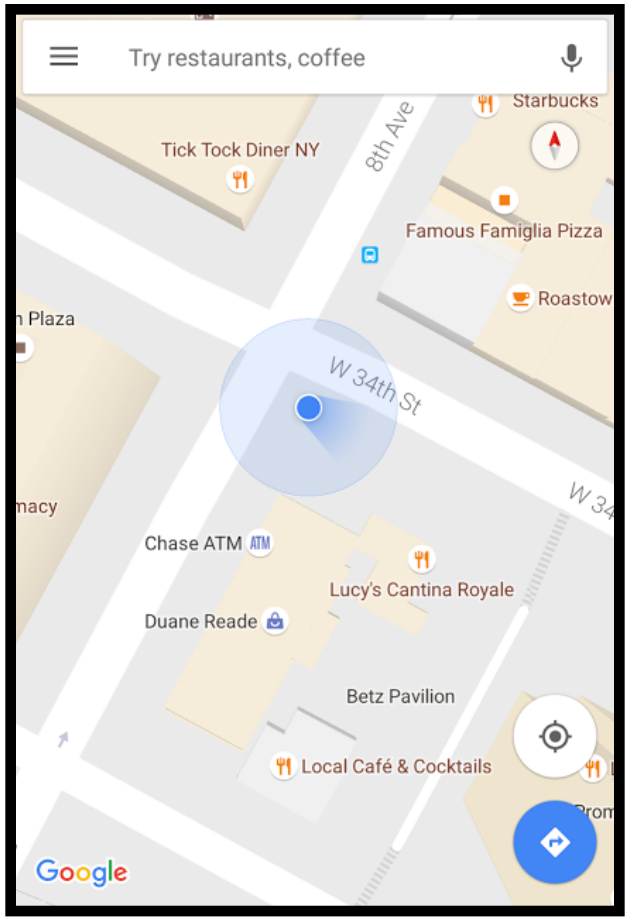
Chandrakana Nandi, Dan Grossman,
Adrian Sampson, Todd Mytkowicz,
Kathryn S. McKinley

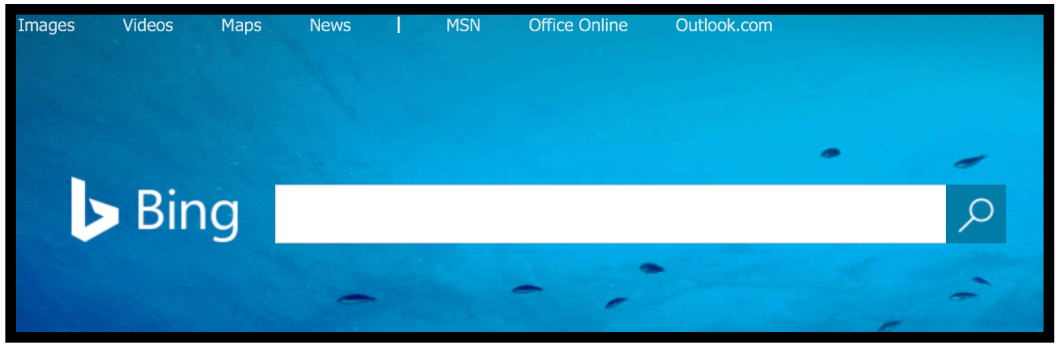


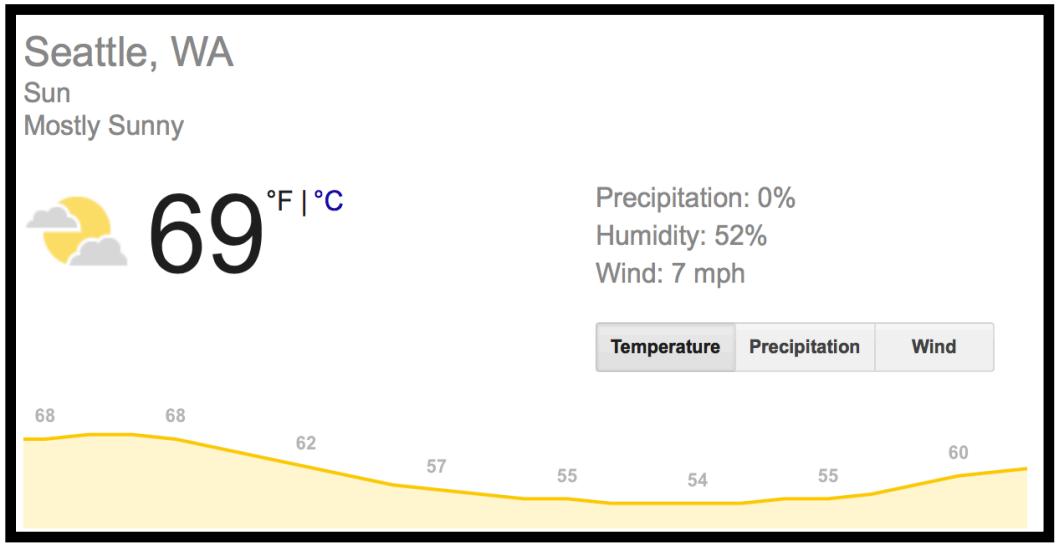
Cornell University

Microsoft® Research

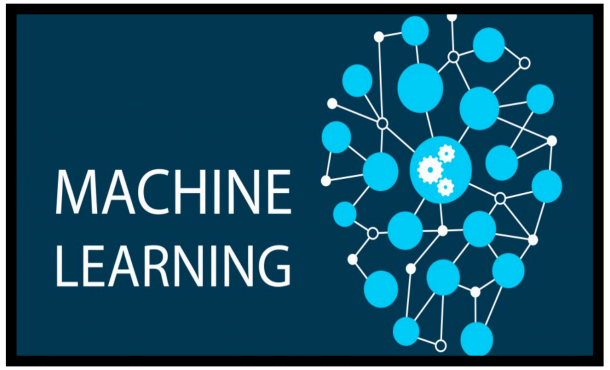
Google

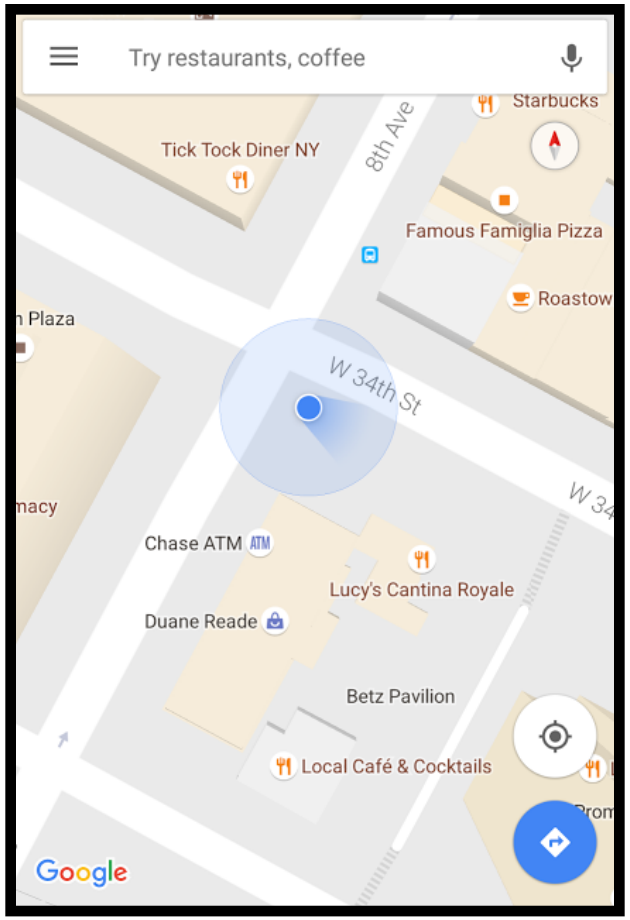

Try restaurants, coffee
Starbucks
Tick Tock Diner NY
8th Ave
Famous Famiglia Pizza
Roastow
Plaza
W 34th St
macy
Chase ATM
Lucy's Cantina Royale
Duane Reade
Betz Pavilion
Local Café & Cocktails
Google

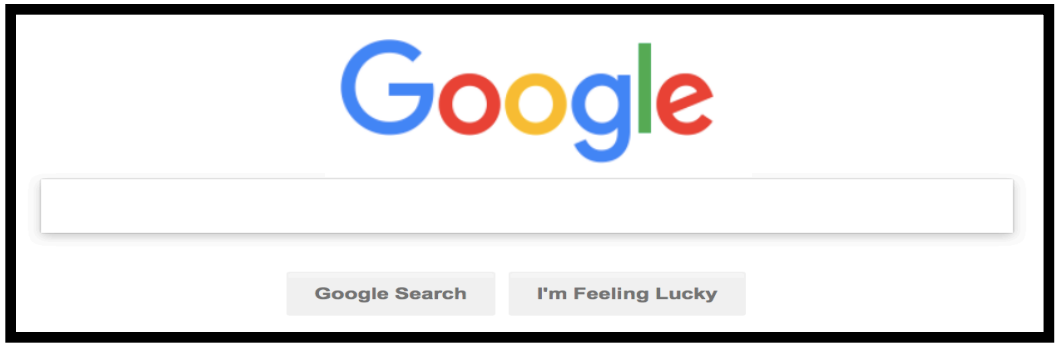

Images
Videos
Maps
News
MSN
Office Online
Outlook.com
Bing

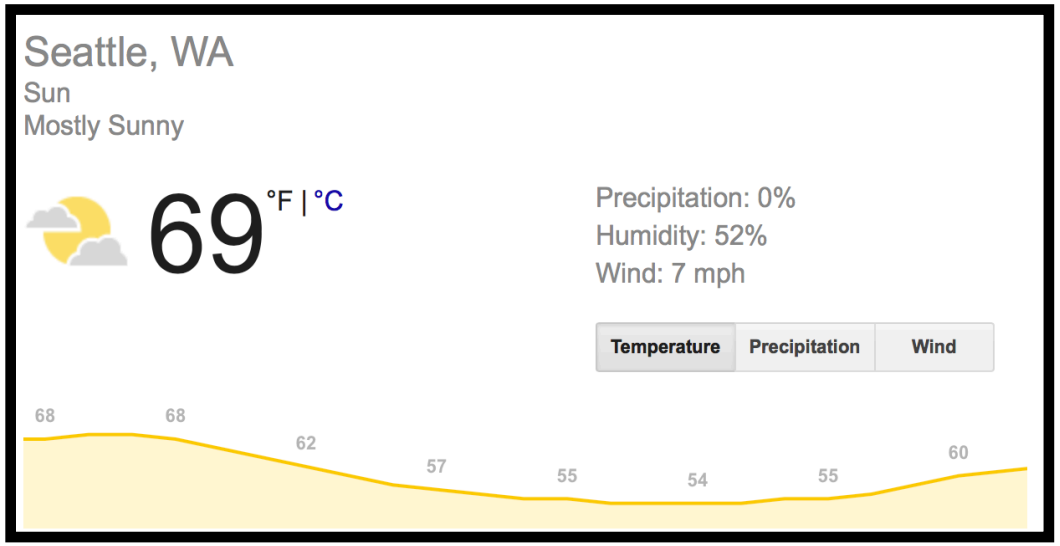

Seattle, WA
Sun
Mostly Sunny
69 °F | °C
Precipitation: 0%
Humidity: 52%
Wind: 7 mph
Temperature
Precipitation
Wind
68
68
62
57
55
54
55
60


SMART HOUSE

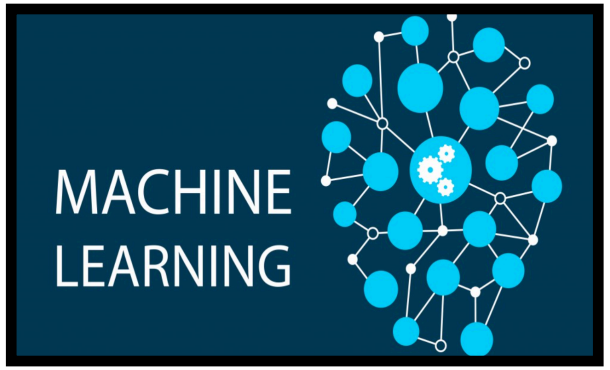

MACHINE
LEARNING


Try restaurants, coffee
Starbucks
Tick Tock Diner NY
8th Ave
Famous Famiglia Pizza
Roastow
Plaza
W 34th St
macy
Chase ATM
Lucy's Cantina Royale
Duane Reade
Betz Pavilion
Local Café & Cocktails
Google


Google
Google Search
I'm Feeling Lucky


Seattle, WA
Sun
Mostly Sunny
69 °F | °C
Precipitation: 0%
Humidity: 52%
Wind: 7 mph
Temperature
Precipitation
Wind
68
68
62
57
55
54
55
60


SMART HOUSE


MACHINE
LEARNING

# What can go wrong in probabilistic programs?

All the things that can always go wrong

Approximation errors

Dependence errors

Model errors

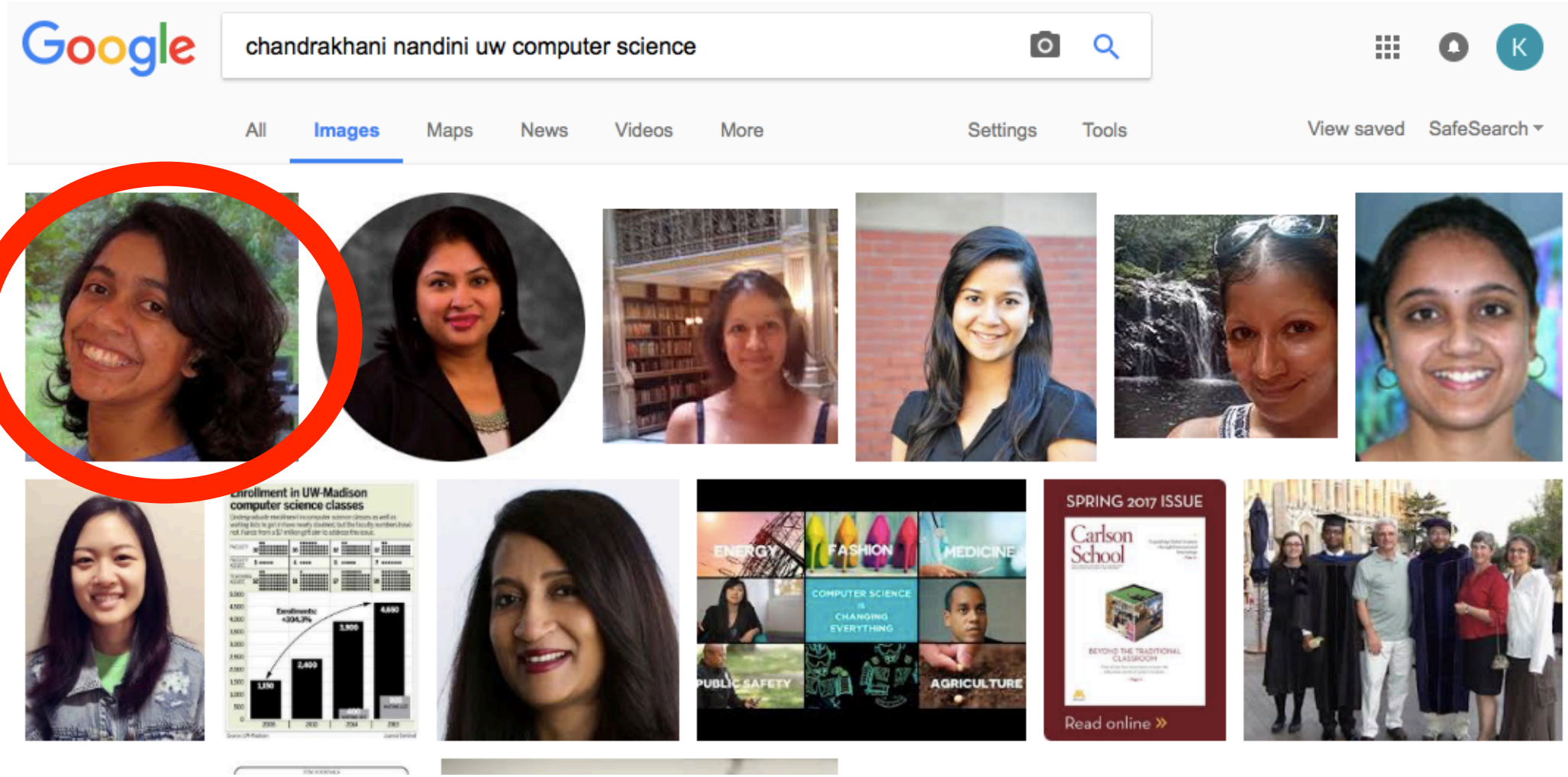
# Add author photos to papers

Image search for author names

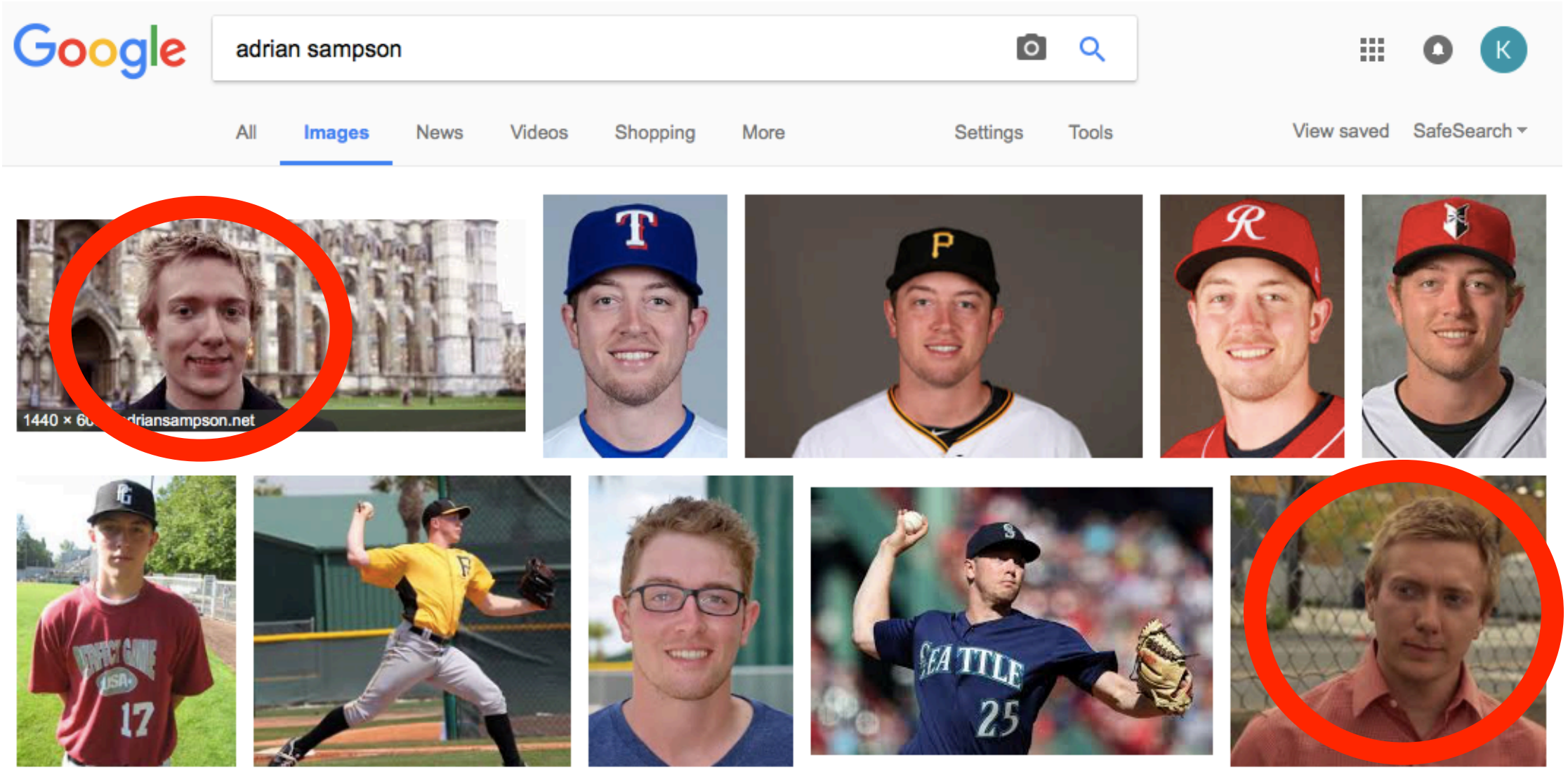
Approximation

Use top (k = 1) ranked image?

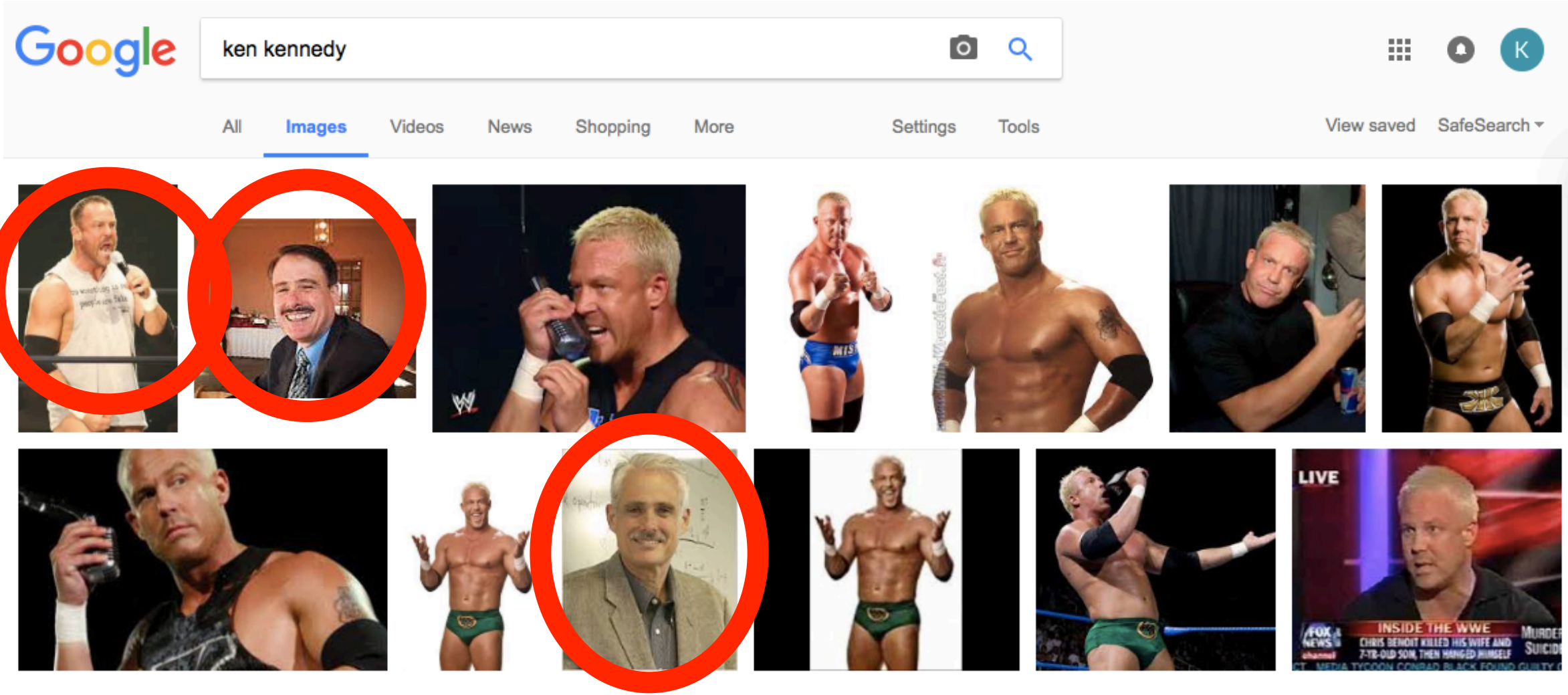
# Add author photos to papers



# Add author photos to papers



# Add author photos to papers

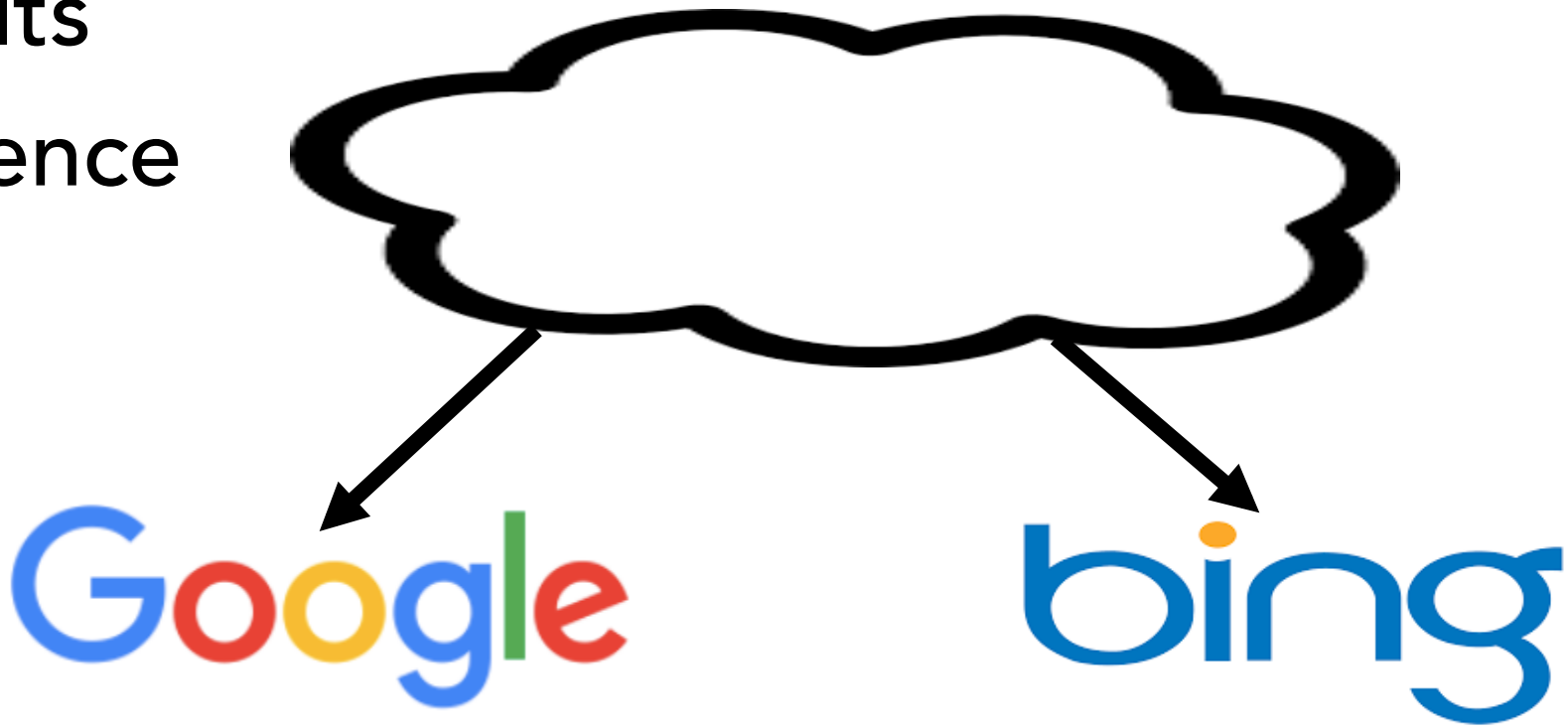


# Add author photos to papers

Search on Google and bing for author names

Combine results

Input dependence



# Statistics for debugging

Approximation error

- Sufficient samples? Correct hyper-parameters?

Dependence errors

- Input correlation? Premature inference calls?

Incorrect models

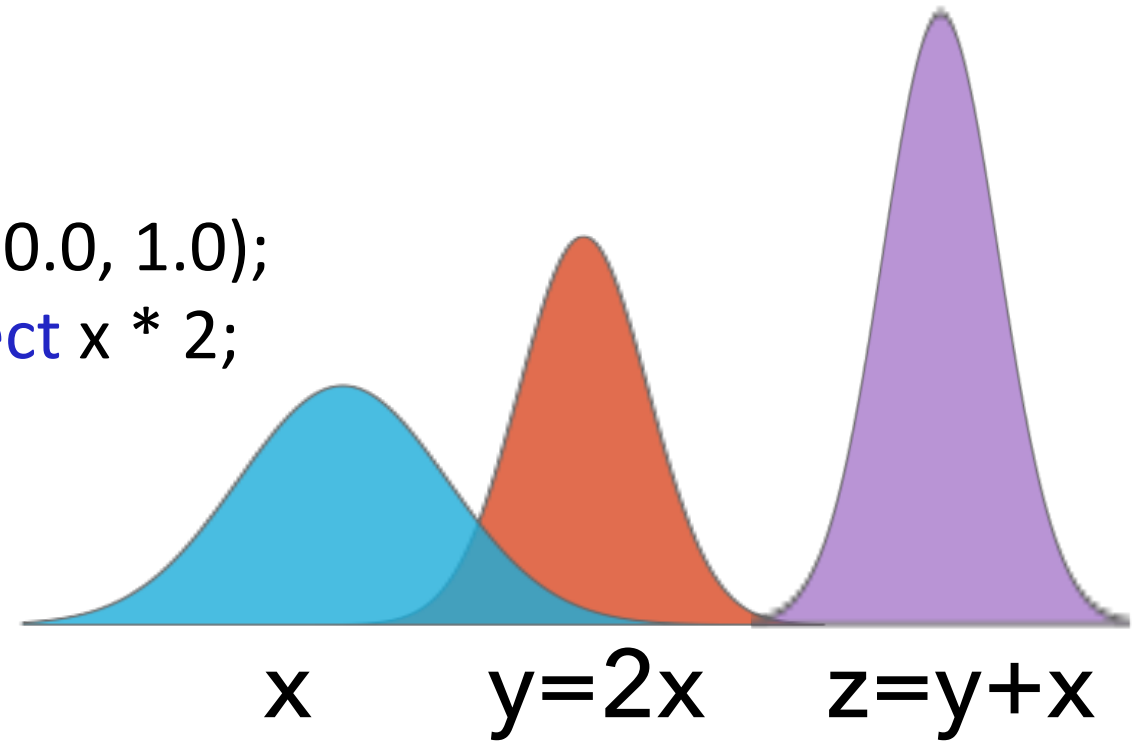
- Are the distributions correct?

# Programming model and representation

Uncertain<T> + explicit inference

# Uncertain <T>

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Z = from y in Y
                      from x in X
                      select y + x;
```

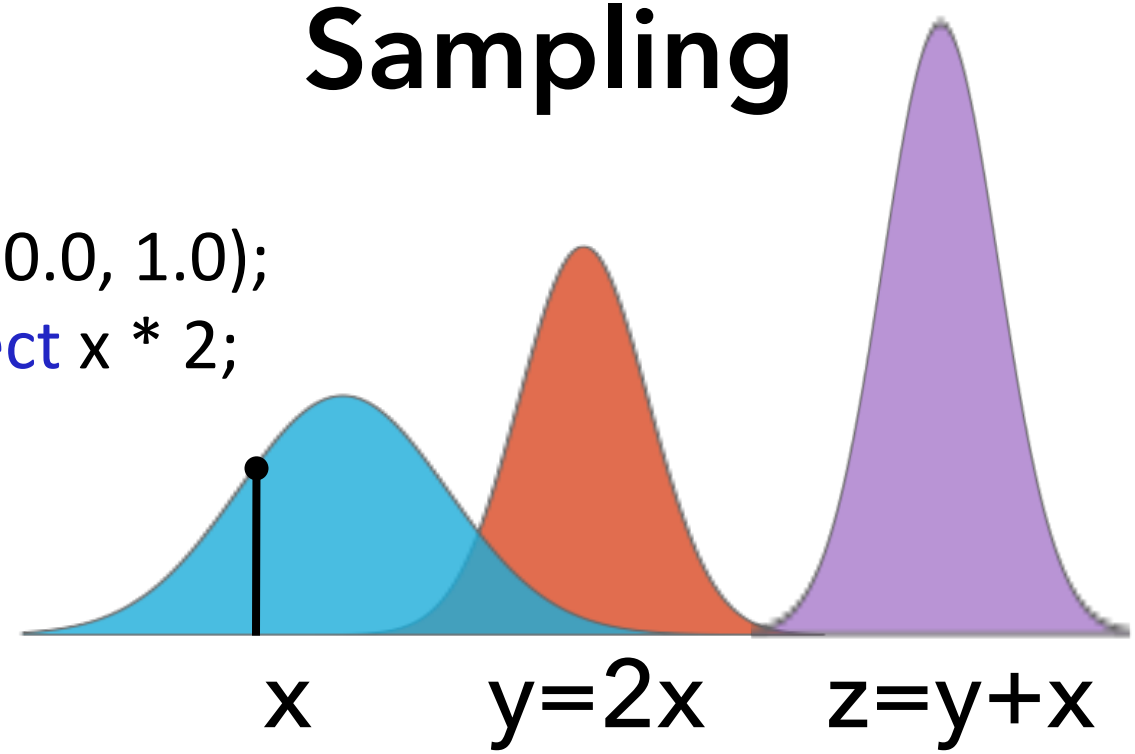


# Uncertain <T>

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Z = from y in Y
                      from x in X
                      select y + x;
if (Z < 6.0) { // Hypothesis test
   print("2-sigma rule holds for Z.");
}
```

# Implicit inference

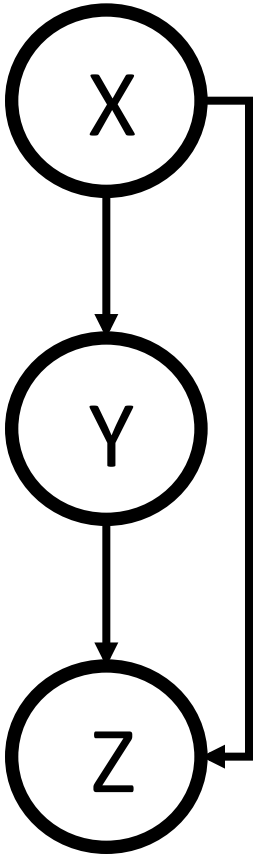
# Sampling



# Uncertain <T>

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Z = from y in Y
                      from x in X
                      select y + x;
if (Z < 6.0) { // Hypothesis test
   print("2-sigma rule holds for Z.");
}
```

# Bayesian Network



# FLEXI: U<T> + explicit inference

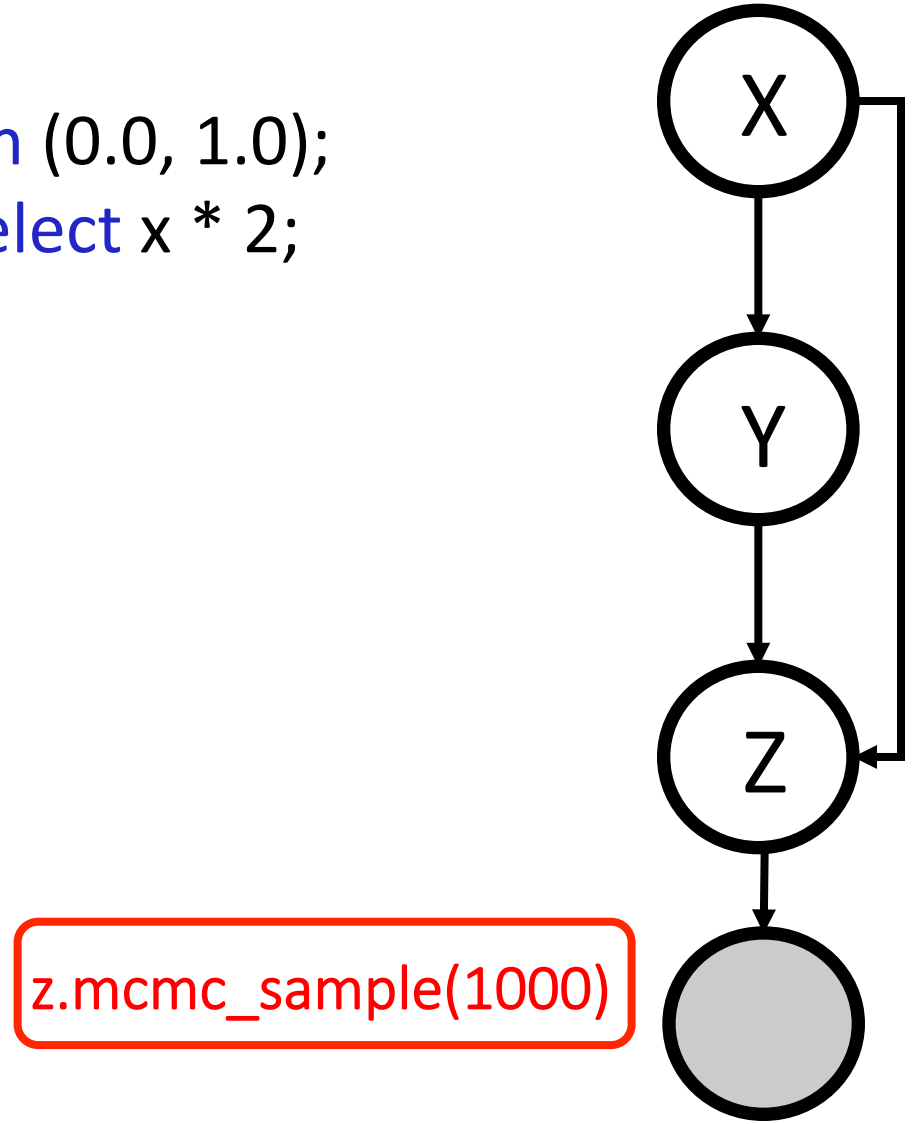
Inference is first class

Programs can produce, compose, and externalize distributions

**Pitfalls:** where to put inference and is it correct?

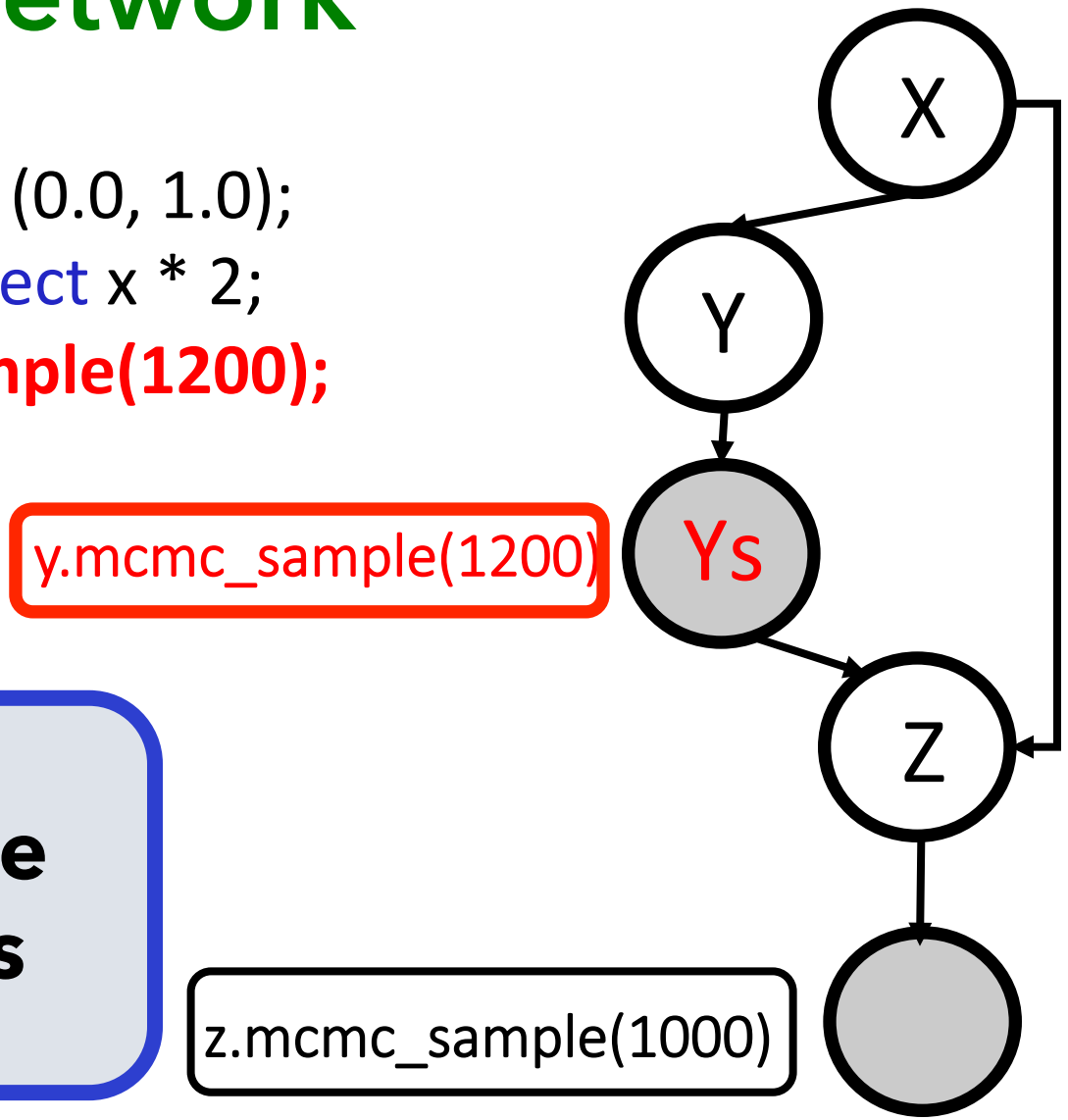
# Decorated Bayesian Network (DBN)

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Z = from y in Y
                      from x in X
                      select y + x;
if (Z < 6.0) {    // Hypothesis test
   print("2-sigma rule holds for Z.");
}
```



# Decorated Bayesian Network

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Ys = Y.mcmc_sample(1200);
Uncertain<double> Z = from y in Ys
                      from x in X
                      select y + x;
```




**Helps identify dependence and approximation errors**

# Statistics for debugging

Approximation error

- Sufficient samples? Correct hyper-parameters?

Dependence errors

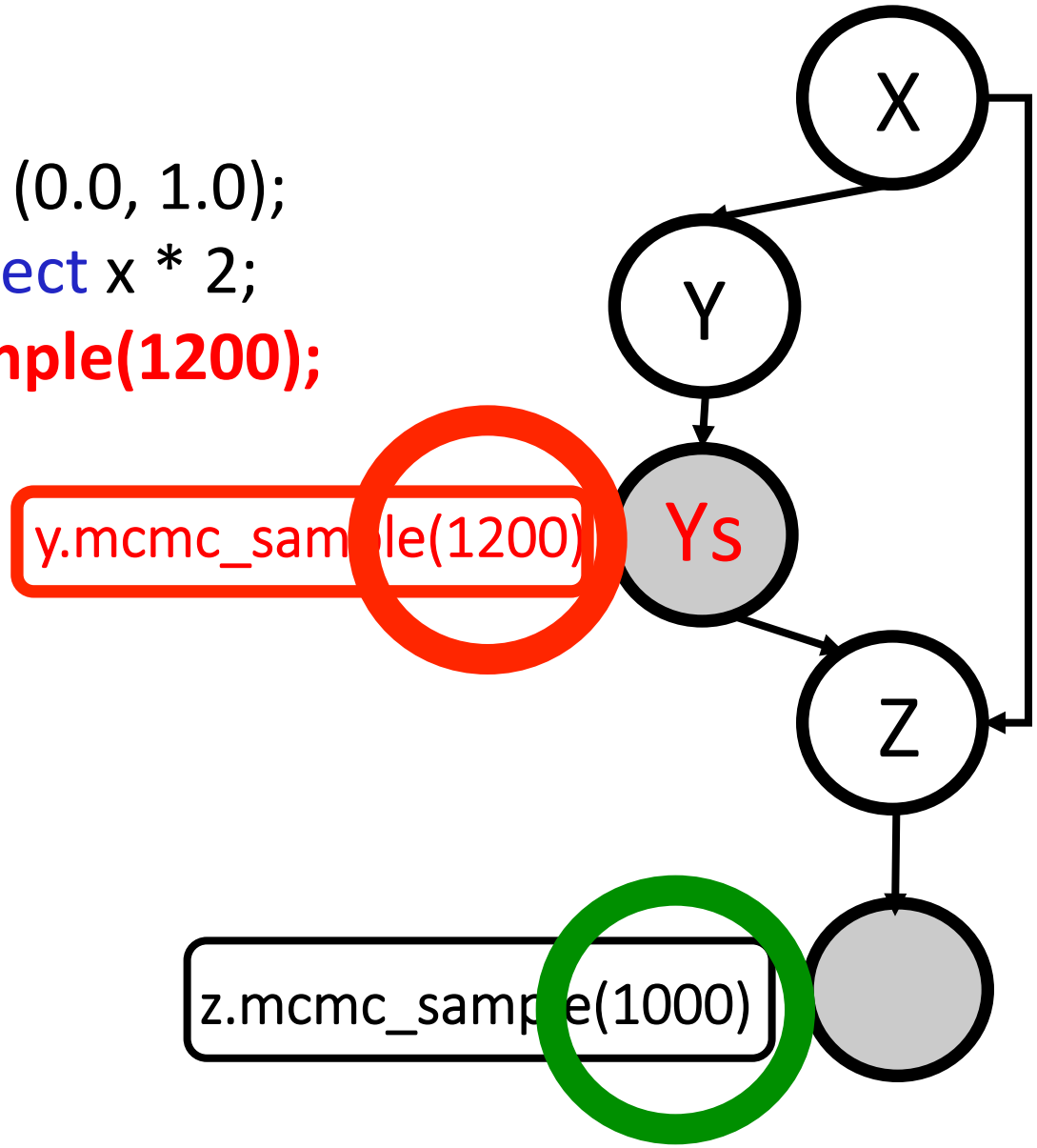
- Input correlation? Premature inference calls?

Incorrect models

- Are the distributions correct?

# Sample size

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Ys = Y.mcmc_sample(1200);
Uncertain<double> Z = from y in Ys
                      from x in X
                      select y + x;
if (Z < 6.0) {    // Hypothesis test
   print("2-sigma rule holds for Z.");
}
```



# Meta inference for hyper-parameters h

Sample size

Learning rate

Top k

# Meta inference for h

***p*** probabilistic program

***h*** hyper-parameter

***c*** domain specific correctness requirement

Pr (*p* is correct) = Pr (*c* is met && *h* is good)

= Pr (*c* is met | *h* is good) * Pr(*h* is good)

Pr(*h* is good): if *h* maximizes its likelihood function

# Meta inference for sample size

*h* = sample size (n)

fix prior over n, f(n | p)

find n as $\text{argmax}_n$ L(p | n) / variance ($s_n$)

where L(p | n) is the likelihood p and

variance ($s_n$) is the variance of the sample

# Evaluation of meta inference (MI)

| Benchmark | h | Time (sec) with MI | Time (sec) without MI | Default value |
|---|---|---|---|---|
| Linear regression | Learning rate | 0.04 | 0.05 | Same as MI |
| Search engine | Top-k | 100.70 | 62.10 | 3 |
| Exponential sampler | Sample size | 17.40 | 0.10 | 1000 |

# Statistics for debugging

Approximation error

- Sufficient samples? Correct hyper-parameters?

Dependence errors

- Input correlation? Premature inference calls?

Incorrect models

- Are the distributions correct?

# Data dependence (correlation) detection

Let $v_1$, $v_2$, … $v_n$ be random variables in the program

Let $r_{ij}$ be the correlation coefficient between $v_i$ and $v_j$

Compute the correlation matrix for all i, j

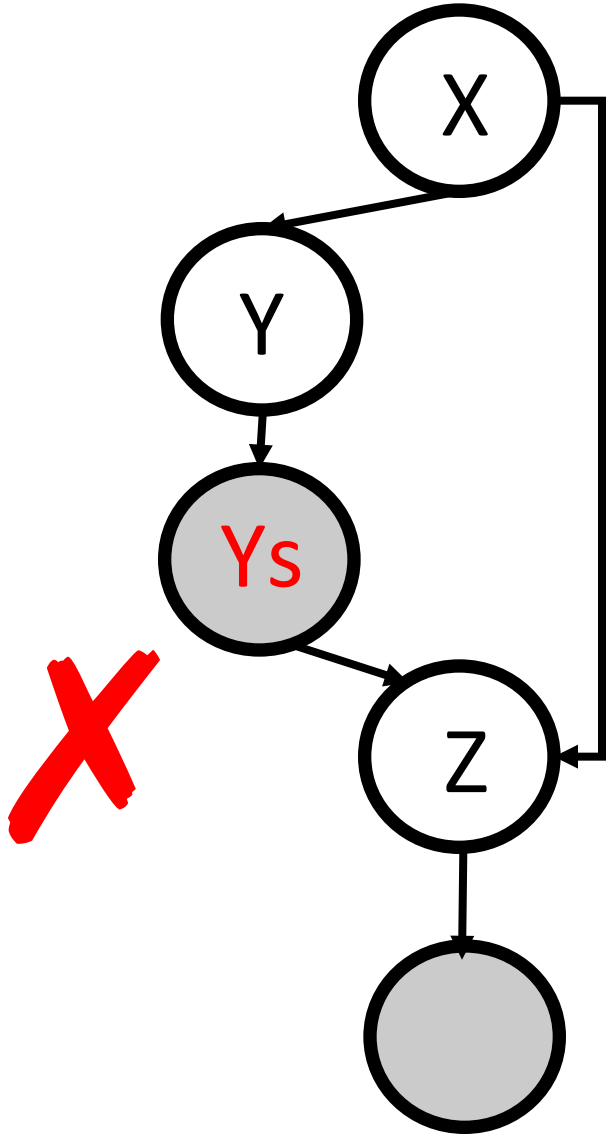
**If $| r_{ij} | > .7$ (default), then $v_i$ and $v_j$ are correlated, but if $v_i$ and $v_j$ are independent in the DBN, it is an error**

# Data dependence (correlation) detection

```
Uncertain<double> X = new read.distribution(temperature.year);
Uncertain<double> Y = new read.distribution(humidity.year);
Uncertain<double> Z = from x in X
                      from y in Y
                      select f(x, y);
...
```

# Premature call to inference

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Ys = Y.mcmc_sample(1200);
Uncertain<double> Z = from y in Ys
                      from x in X
                      select y + x;
if (Z < 6.0) {    // Hypothesis test
   print("2-sigma rule holds for Z.");
}
```



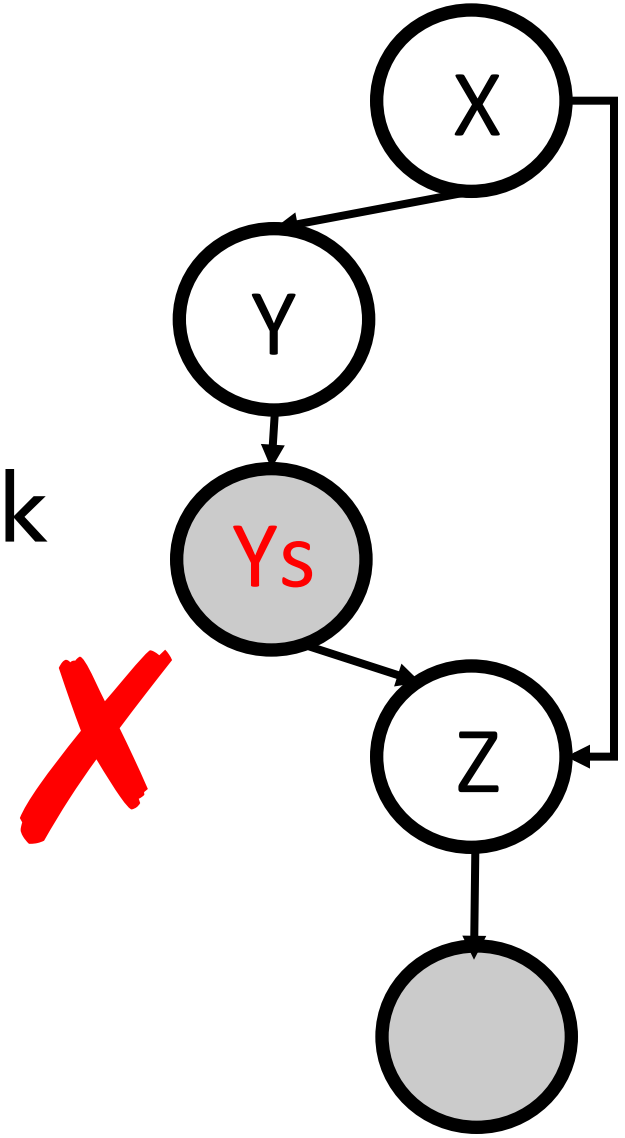
# Premature call to inference

Given an inference node i in the DBN

For all descendants k of i

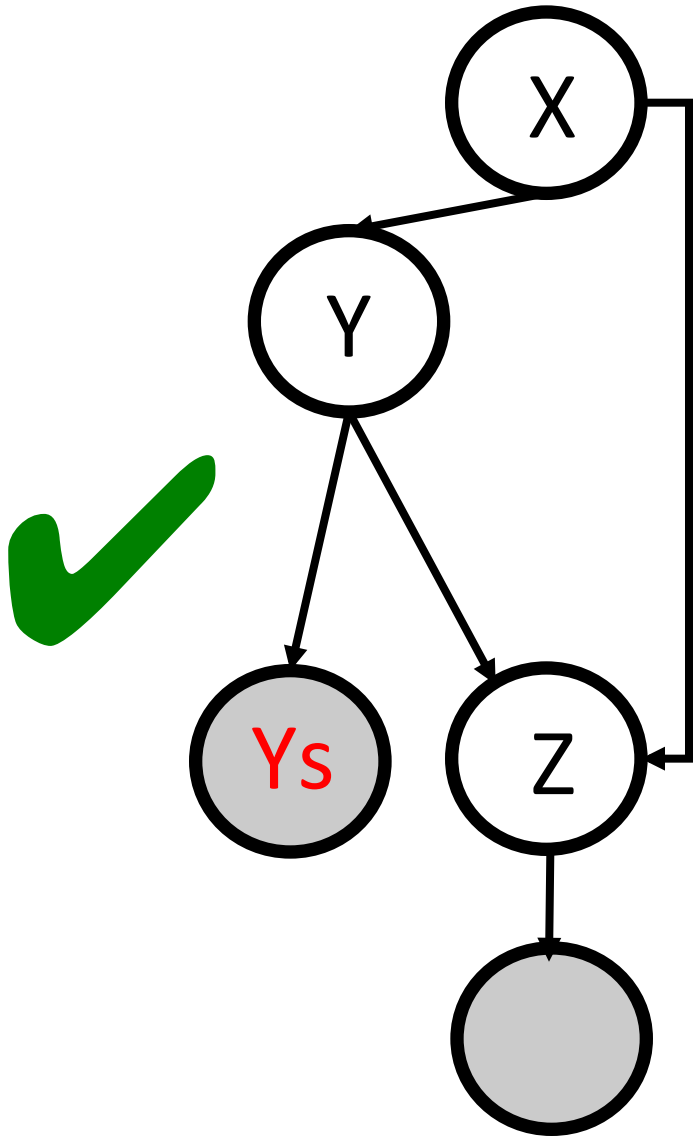
if E a path from an ancestor of i to k

then inference is premature



# Correct call to inference

```
Uncertain<double> X = new Gaussian (0.0, 1.0);
Uncertain<double> Y = from x in X select x * 2;
Uncertain<double> Z = from y in Y
                      from x in X
                      select y + x;
if (Z < 6.0) {    // Hypothesis test
   print("2-sigma rule holds for Z.");
}
Uncertain<double> Ys = Y.mcmc_sample(1200);
```



# Statistics for debugging

Approximation error

- Sufficient samples? Correct hyper-parameters?

Dependence errors

- Input correlation? Premature inference calls?

Incorrect models

- Are the distributions correct?

# Incorrect models

What is the program suppose to do? [PLDI 2014]

`passert e, p, c`

*e* must hold with probability *p*
at confidence *c*

# Probabilistic assertion

```
true_avg = average(salaries)
private_avg =
  average(obfuscate(salaries))
passert (true_avg - private_avg
        <= 10,000), 90, 99
```

# Statistics for debugging

Sampling based approach

If the passert fails or the inference does not

terminate for large n

then DePP reports

poor data or model errors are likely

# Probabilistic programming is in its infancy

Identify 3 classes of new statistical bugs

New DBN program representation

Decorated Bayesian Networks

DePP debugger shows potential