ASSESSING PRIVACY RISKS USING PROBABILISTIC PROGRAMMING

Unitec Institute of Technology TE WHARE WĀNANGA O WAIRAKA

IT UNIVERSITY OF COPENHAGEN

Raúl Pardo (raup@itu.dk), Willard Rafnsson (wilr@itu.dk), Christian Probst (cprobst@unitec.ac.nz) and Andrzej Wąsowski (wasowski@itu.dk)

PROBLEM

Attacker knowledge

Name	Age
Alice	$\mathcal{U}(0,100)$
Bob	$\mathcal{C}(55.2)$
Charlie	$\mathcal{C}(55.2)$
Dave	$\mathcal{C}(55.2)$

knows-a-lot (*Kalo*)

Name	Age
Alice	$\mathcal{U}(0,100)$
Bob	$\mathcal{N}(55.2,3.5)$
Charlie	$\mathcal{N}(55.2,3.5)$
Dave	$\mathcal{N}(55.2,3.5)$

knows-a-little (*Kali*)

- What can the attackers (*Kalo* and *Kali*) say about Alice´s age (a) after observing that the output of the program (o) is 55.3?
- What attacker learned more about Alice´s age by observing the output of the program?

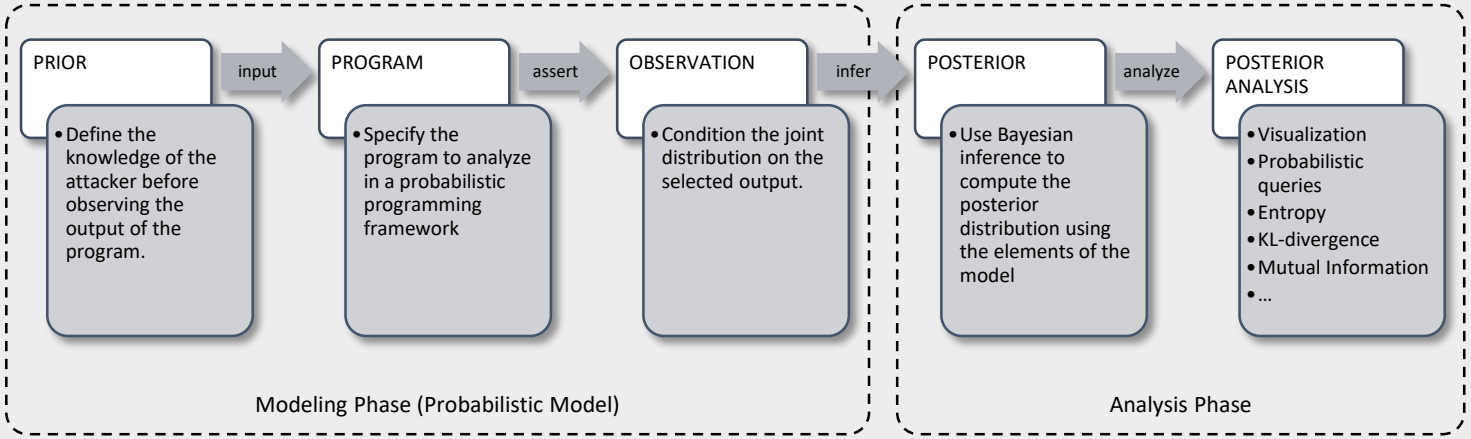
Program

```
def agg(records: List[(String,Double)]): Double =
  records.map    { case(n,a) => (a,1) }
         .reduce { case(x,y) => (x._1+y._1,x._2+y._2) }
         .map    { case(sum,count) => sum/count}
```

Observed output

55.3

METHOD



PRIVACY RISK ANALYSES

Quantiative Information Flow metrics and other statistics

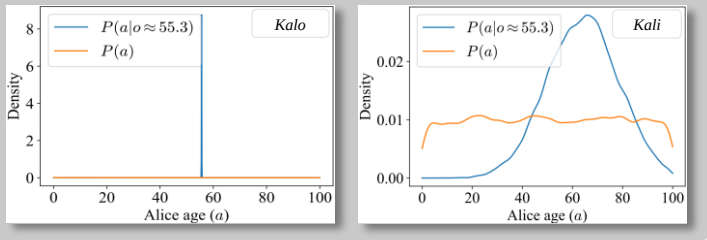
		Kalo	Kali
Expectation/ Standard deviation	$\mathbb{E}[a\|o \approx 55.3] \pm \sigma[a\|o \approx 55.3]$	55.6 ± 0.01	64 ± 14
Probability query	$P(a \leq 18\|o \approx 55.3)$	0	0.004
Shannon Entropy	$H(a\|o \approx 55.3)$	-3.08	5.83
KL-divergence	$D_{KL}(a\|o \approx 55.3\|\|a)$	5.64	0.77
Mutual Information	$I(a;o)$	9.37	0.60

(any metrics derived from MCMC chains can easily be added)

Prior/Posterior knowledge about Alice´s age. Kalo ‘s posterior reduces to a point distribution (no uncertainty). Kali´s posterior still has a large standard deviation (some uncertainty).

- Kalo´s knowledge about Alice´s age is more precise than Kali´s
- Both Kalo and Kali learned that Alice is not underage
- Kalo´s uncertainty about Alice´s age is much lower than Kali's
- Kalo´s knowledge has changed more than Kali´s after observing the ouput
- The output contains more information about Alice´s age for Kalo than for Kali

KDE plots and other visualizations



Probabilistic programming *can be effectively used to* ***perform*** *a wide range of* ***quantitative analyses*** *to find* ***privacy leaks*** *in data science programs.*