Hierarchical models in Stan

QoCT surveys

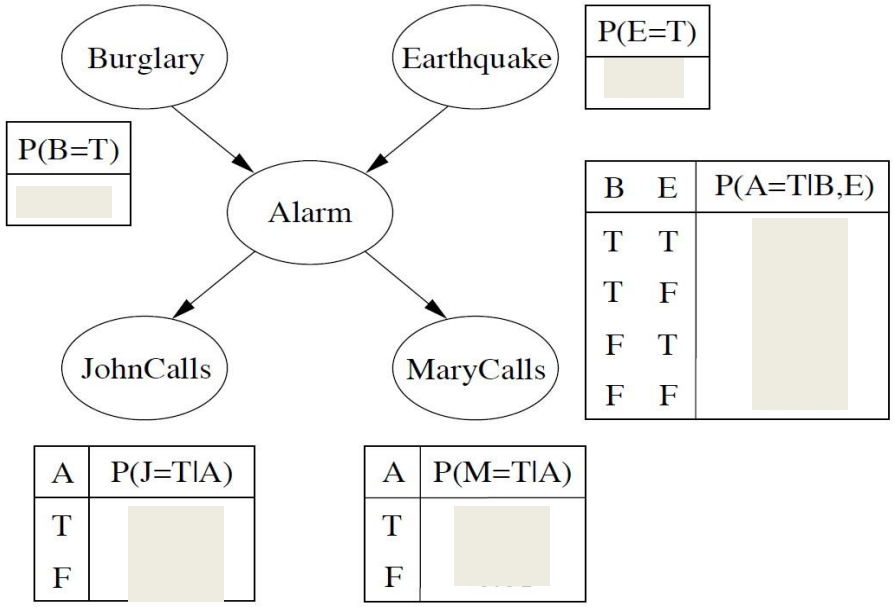
The link is at: https://apps.eng.unimelb.edu.au/casmas/index.php?r=qoct/subjects.

Workshop # 11

Prepared by: Yasmeen George

Bayesian Network: Training

- Constructing the structure of the network
  - domain expert to decide the causal relations
  - *structure learning* algorithms exist, but complicated
- Parameter learning (filling the table)

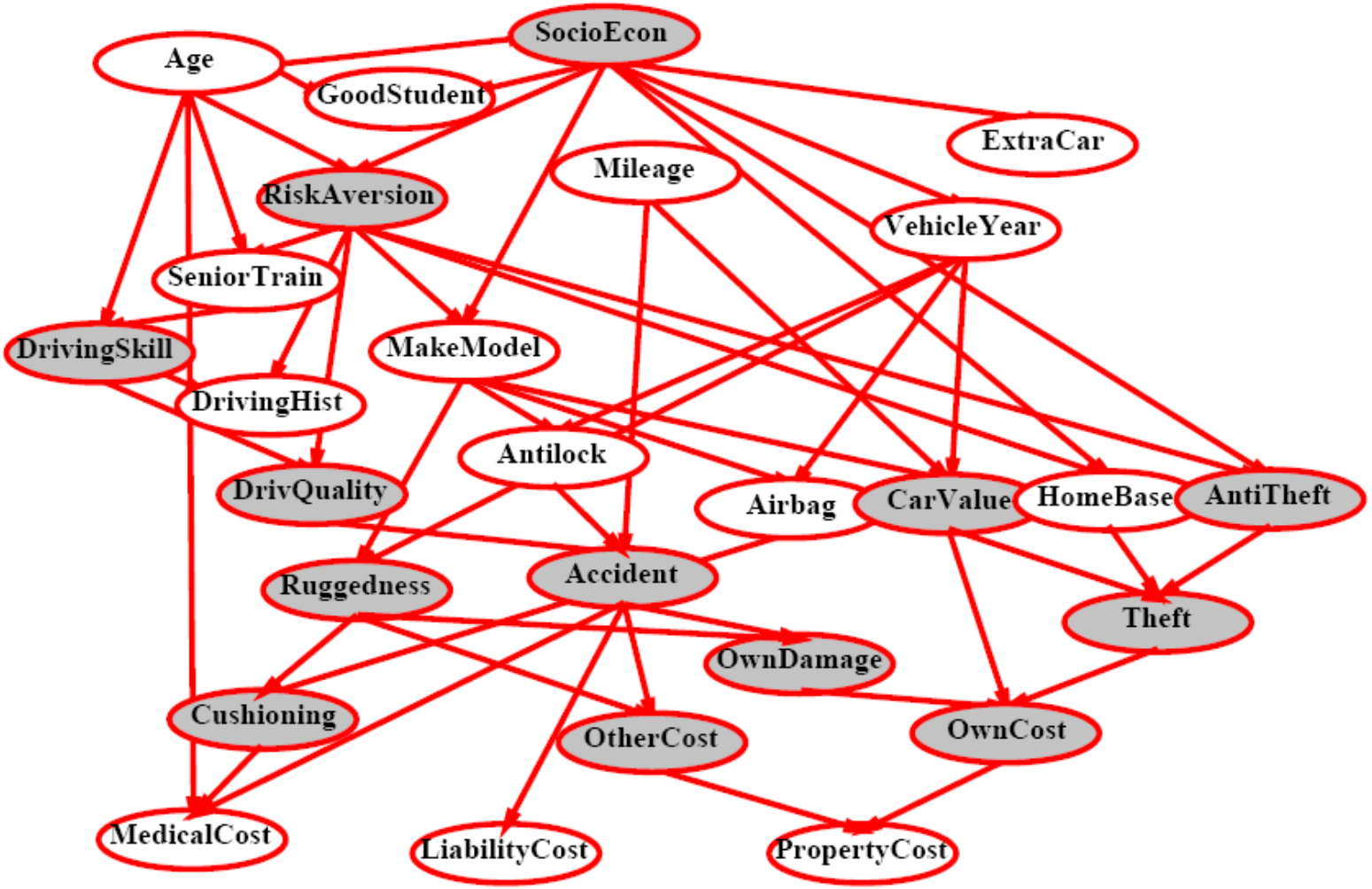


Training

A	B	E	J	M
T	F	T	F	T
F	T	F	F	F
T	F	T	F	T
…				

  - Using EM method if there are missing values

A more realistic Bayes Network: Car insurance



in reality, parameter estimation is not an easy task !!

Motivation for Stan

- Fit rich Bayesian statistical models

- The Process
    1. Create a statistical model
    2. Perform inference on the model
    3. Evaluate

What is Stan?

- Statistical model specification language
- 4 cross-platform users interfaces
    - CmdStan - command line
    - RStan - R integration
    - PyStan - Python integration
    - MStan - Matlab integration (user contributed)

The Stan Language

- Data Types
  - basic: `real`, `int`, `vector`, `row_vector`, `matrix`
  - constrained: `simplex`, `unit_vector`, `ordered`, `positive_ordered`, `corr_matrix`, `cov_matrix`
  - arrays
- Bounded variables
  - applies to `int`, `real`, and matrix types
  - lower example: `real<lower=0> sigma;`
  - upper example: `real<upper=100> x;`
- Program Blocks
  - data (optional)
  - transformed data (optional)
  - parameters (optional)
  - transformed parameters (optional)
  - model
  - generated quantities (optional)

Stan Example: vectorization

```
data {
  int<lower=0> N;
  vector[N] y;
  vector[N] x;
}
parameters {
  real alpha;
  real beta;
  real<lower=0> sigma;
}
model {
  alpha ~ normal(0,10);
  beta ~ normal(0,10);
  sigma ~ cauchy(0,5);
  y ~ normal(alpha + beta * x, sigma);
}
```

what is this model ????

RStan Demo 1

Eight Schools: hierarchical example

- Educational Testing Service study to analyse the effects of special coaching programs on SAT-V scores in 8 high schools.
- The observed effects of special preparation are estimates based on separate analyses for the eight school experiments.
- The effects, are labelled as Yj . Over 30 students were tested on each school.
- No prior reason to believe any program was:
  - more effective than the others
  - more similar to others

[Rubin, 1981; Gelman et al., *Bayesian Data Analysis*, 2003]

Stan: Eight Schools Data

School	Estimated Treatment Effect	Standard Error of Treatment Effect
A	28	15
B	8	10
C	-3	16
D	7	11
E	-1	9
F	1	11
G	18	10
H	12	18

Eight Schools: Statistical Model

- Estimate hyperparameters ????

```
data {
  int<lower=0> J;           // # schools
  real y[J];                // estimated treatment
  real<lower=0> sigma[J];   // std err of effect
}
parameters {
   real mu;                 // school effect
   real<lower=0> tau;       // mean for schools
   real eta[J];             // variance between schools
}
transformed parameters {
  real theta[J];
    for (j in 1:J)
      theta[j] <- mu + tau * eta[j];
}
model {
  eta ~ normal(0, 1);
  y ~ normal(theta, sigma);
}
```

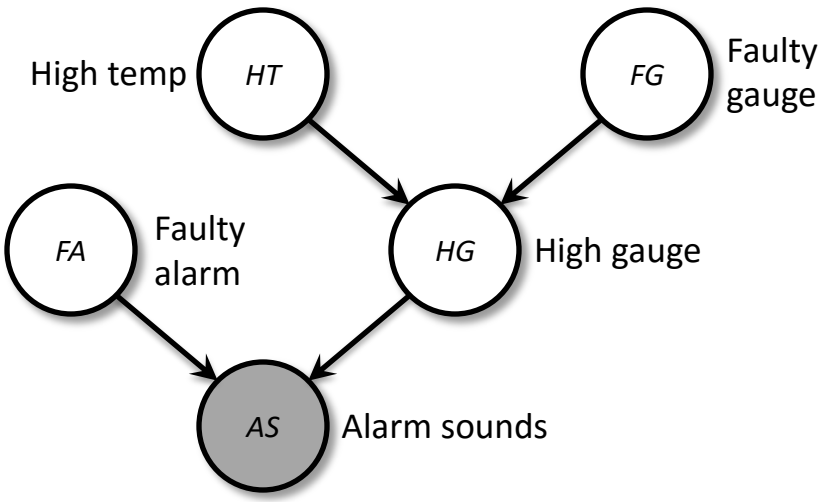
RStan Demo 2

Example: Probability Inference

In a nuclear power station, there is an alarm that senses when a temperature gauge exceeds a given threshold.

The gauge measures the temperature of the core.

Consider the Boolean variables A (alarm sounds), FA (alarm is faulty), and FG (gauge is faulty) and the multivalued, discrete nodes G (gauge reading) and T (actual core temperature).



RStan Demo 3

What is the data block for this model ??

How many parameters ?

What about prior distribution and likelihood ?

Reference

[1] Stan slides: http://astrostatistics.psu.edu/su14/ (Daniel-Lee-Stan-2.pdf)

[2] Hierachical models slides: UIUC Artificial Intelligence (CS440/ECE448)

Stan: Help

- User Guide:
  http://mc-stan.org/manual.html
- Homepage:
  http://mc-stan.org
- Stan Users group:
  https://groups.google.com/d/forum/stan-users