Sampling Based Inference

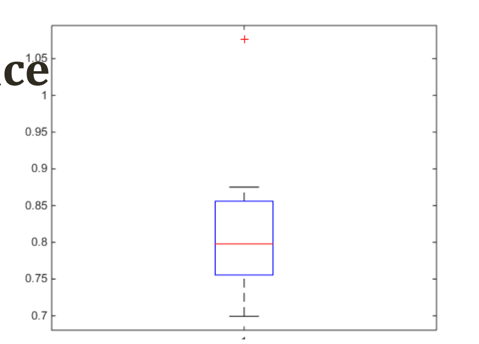
Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

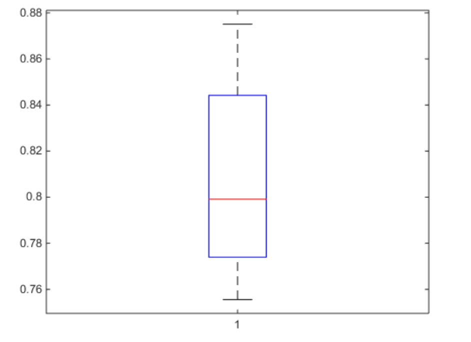
- Learn basic sampling methods
  - Understand the concept of Markov chain Monte Carlo
  - Able to apply MCMC to the parameter inference of Bayesian networks
  - Know the mechanism of rejection sampling
  - Know the mechanism of importance sampling
- Learn sampling based inference
  - Understand the concept of Metropolis-Hastings algorithm
  - Know the mechanism of Gibbs sampling
- Know a case study of sampling based inference
  - Understand the latent Dirichlet allocation model
  - Know the collapsed Gibbs sampling
  - Know how to derive Gibbs sampling formula for LDA

Importance Sampling

Importance Sampling Prone to Extreme Values



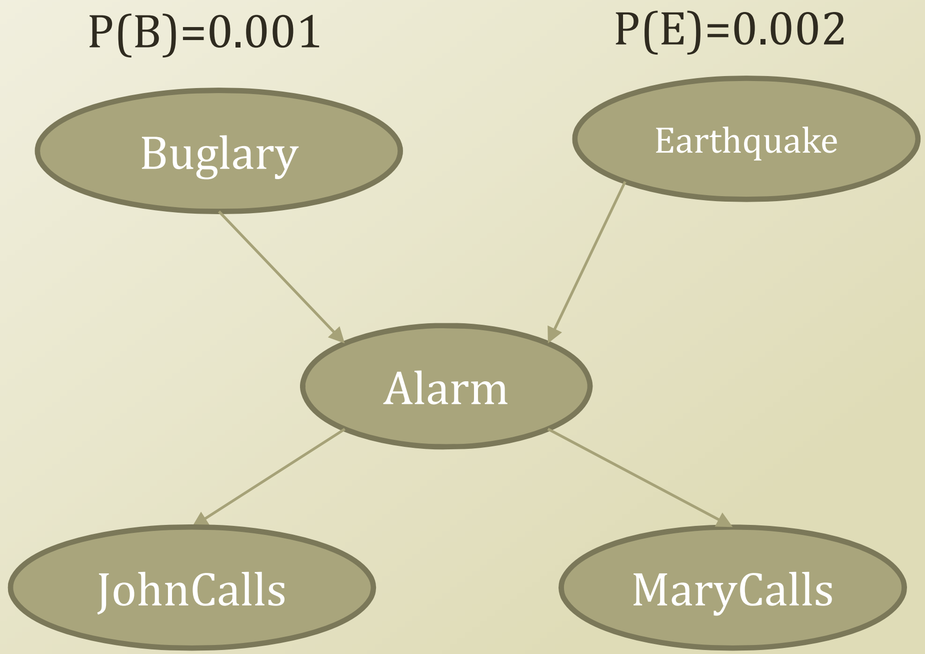
Filtered Extreme Values



- Huge waste from the rejection
- Is generating the PDF the end goal?
  - No… Usually, the question follows
    - Calculating the expectation of PDF
    - Calculating a certain probability
- Let's use the wasted sample to answer the questions
  - $E(f) = \int f(z)p(z)dz = \int f(z)\frac{p(z)}{q(z)}q(z)dz \cong \frac{1}{L}\sum_{l=1}^{L}\frac{P(z^l)}{q(z^l)}f(z^l)$
    - L = # of samples, z^l=a sample of Z
    - Here, the importance weight plays the role
      - $r^l = \frac{P(z^l)}{q(z^l)}$
    - What if $P(z^l)$ and $q(z^l)$ is not normalized, as they should be as probability distributions
    - $E(f) \cong \frac{1}{L}\sum_{l=1}^{L}\frac{P(z^l)}{q(z^l)}f(z^l) = \frac{1}{L}\frac{Z_q}{Z_p}\sum_{l=1}^{L}\frac{\tilde{P}(z^l)}{\tilde{q}(z^l)}f(z^l)$
  - $\mathrm{P(Z>1)} = \int_1^{\infty} 1_{z>1}p(z)dz = \int_1^{\infty} 1_{z>1}\frac{p(z)}{q(z)}q(z)dz \cong \frac{1}{L}\sum_{l=1}^{L}\frac{P(z^l)}{q(z^l)}1_{z^l>1}$

Likelihood Weighting Algorithm

- P(E=T|MC=T,A=F)=?
- LikelihoodWeighting
  - SumSW=NormSW=0
  - Iterate many times
    - SW=SampleWeight = 1
    - Generate a sample from the Bayesian network
      - Buglary → false
      - Earthquake → false
      - Alarm=F|B=F,E=F
        - P(A=F|B=F,E=F)=0.999
        - SW=1*0.999
      - JC|A=T→true
      - MC=T|A=F
        - P(MC=T|A=F)=0.01
        - SW=1*0.999*0.01
    - If the sample has E=T, then SumSW+=SW
    - NormSW+=SW
  - Return SumSW/NormSW
- Any further improvement?
- These samples are....



B	E	P(A|B,E)
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

A	P(J|A)
T	0.90
F	0.05

A	P(M|A)
T	0.70
F	0.01