Machine Learning for Data Science (CS4786)
Lecture 24

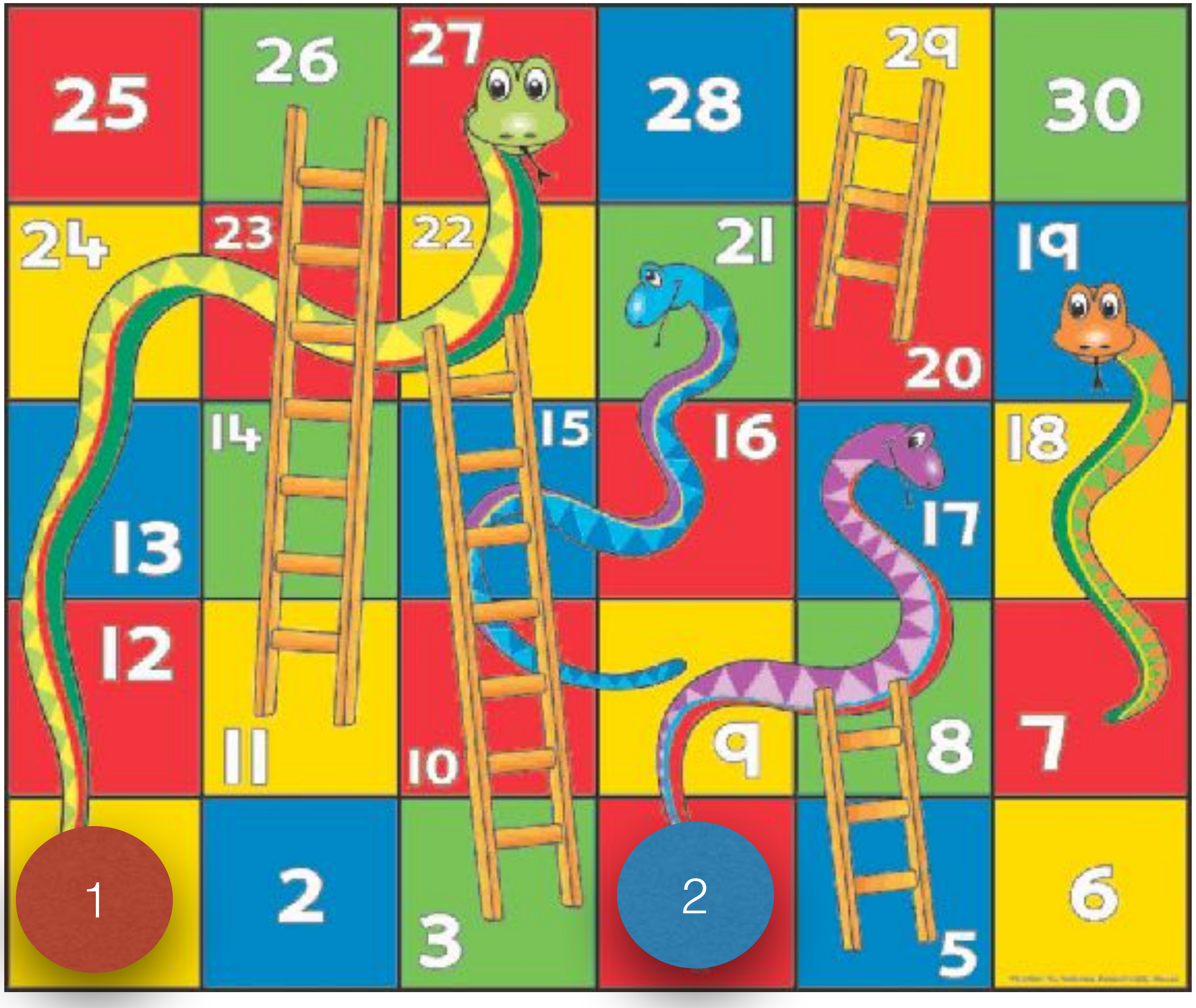
Approximate Inference

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2016fa/

Announcement

- CS Colloquium today at 4:15pm Gates G01 by **Dan Spielman**
    - **“**Laplacian matrices of graphs: Algorithms and Applications”

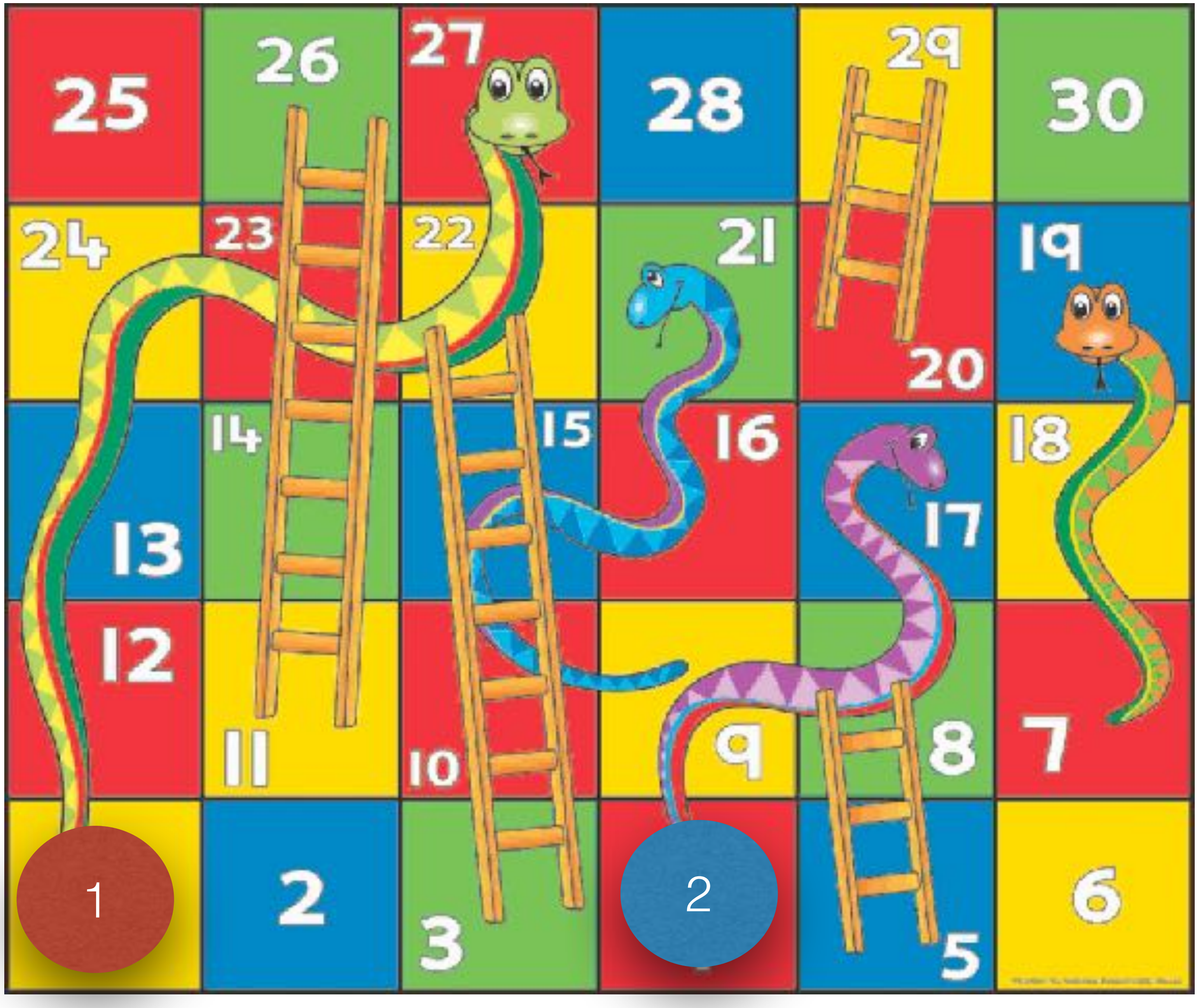
Inference



Who is more likely to win the game?

Compute sum of exact probabilities of all possible sequence of moves leading to Player 1's victory

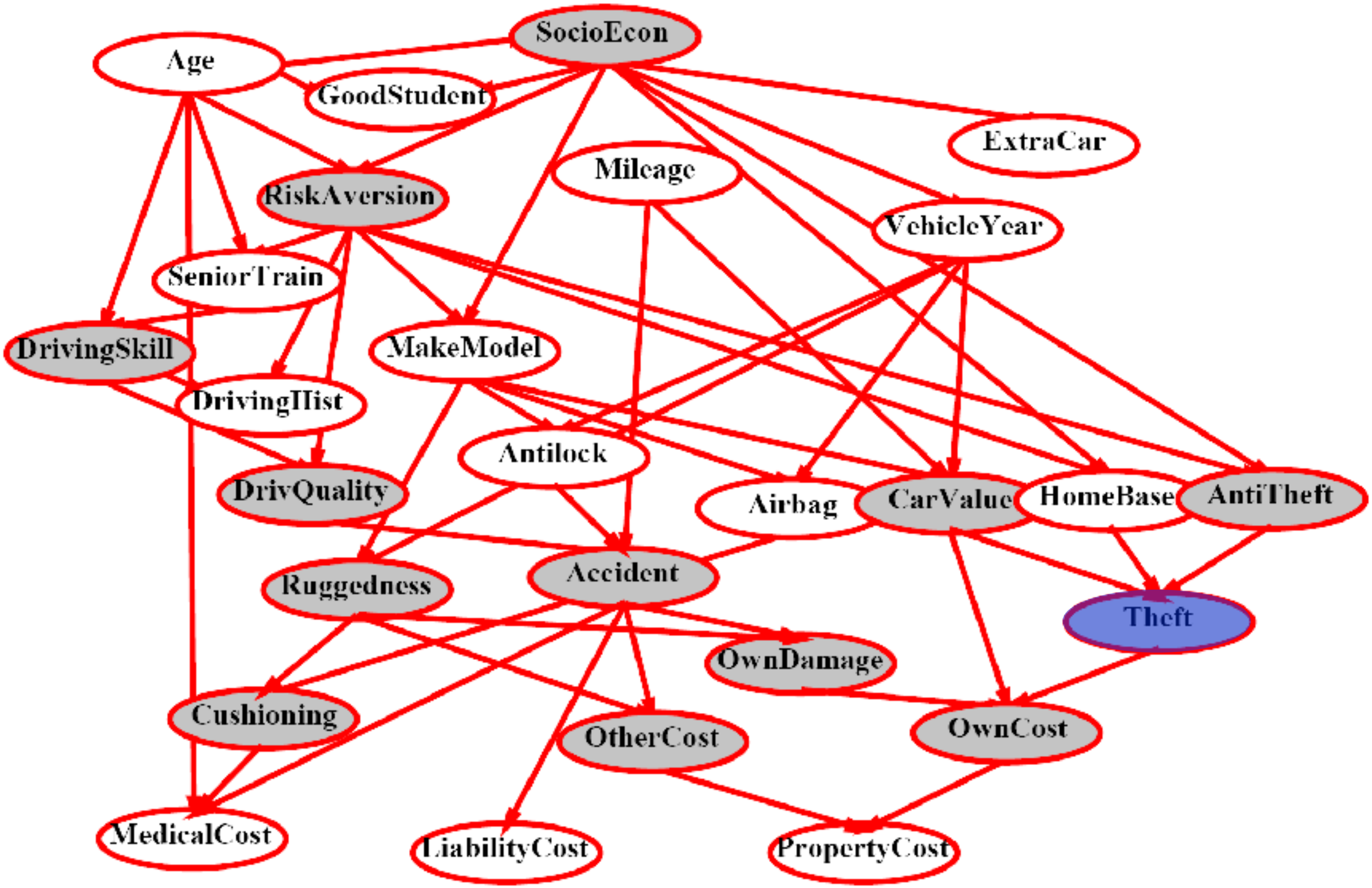
INFERENCE



Who is more likely to win the game?

Throw dice and simulate multiple games, see who wins more often

Inference Via Sampling



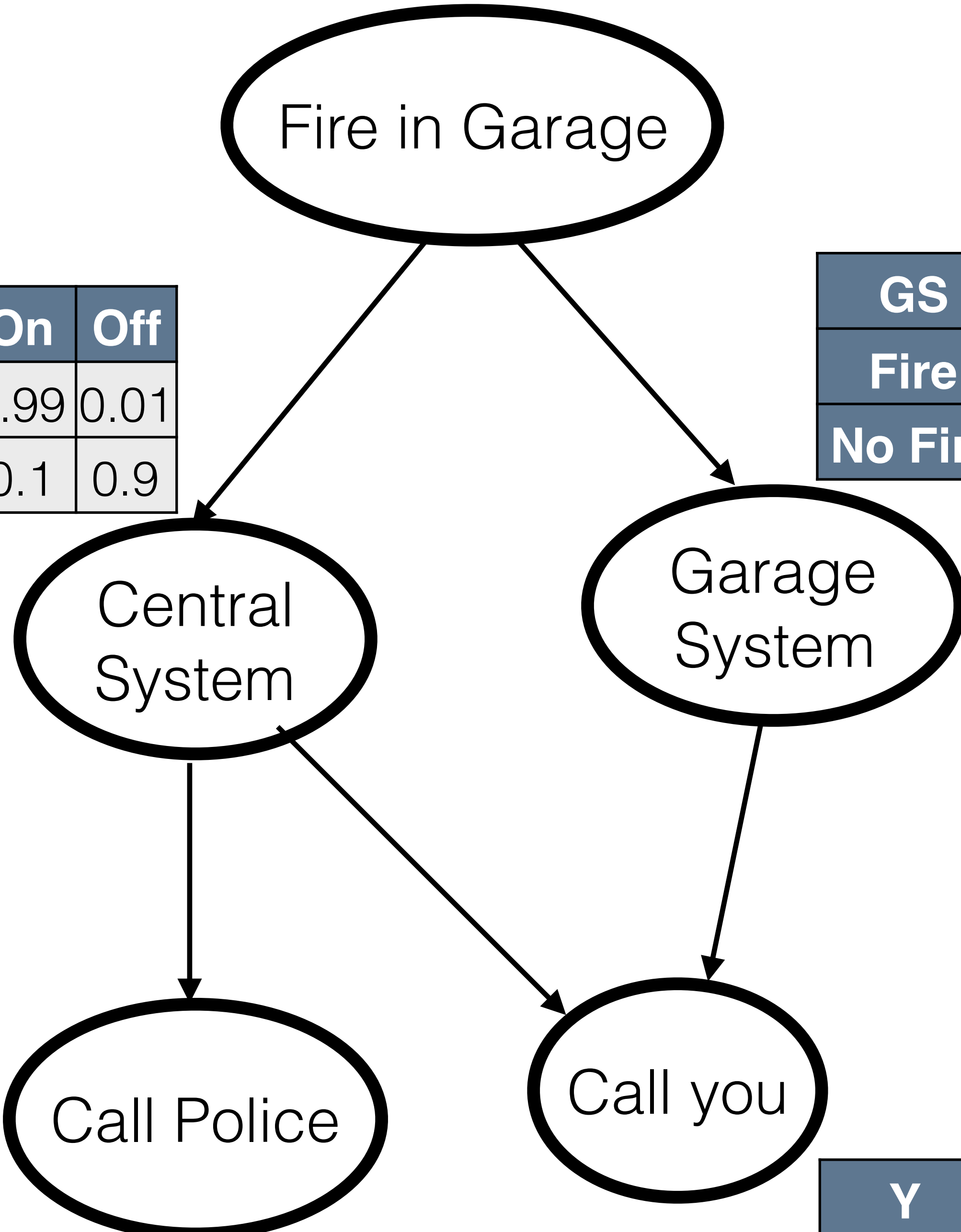
Inference Via Sampling

- Draw n samples from the sampling distribution
- Compute approximate probabilities by computing empirical frequencies
- Why sampling?
  - Getting multiple samples often faster than computing exact probabilities (inference is hard)
  - Inference is key step in learning

REJECTION SAMPLING

	F	CS	GS	P	Y
1					

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

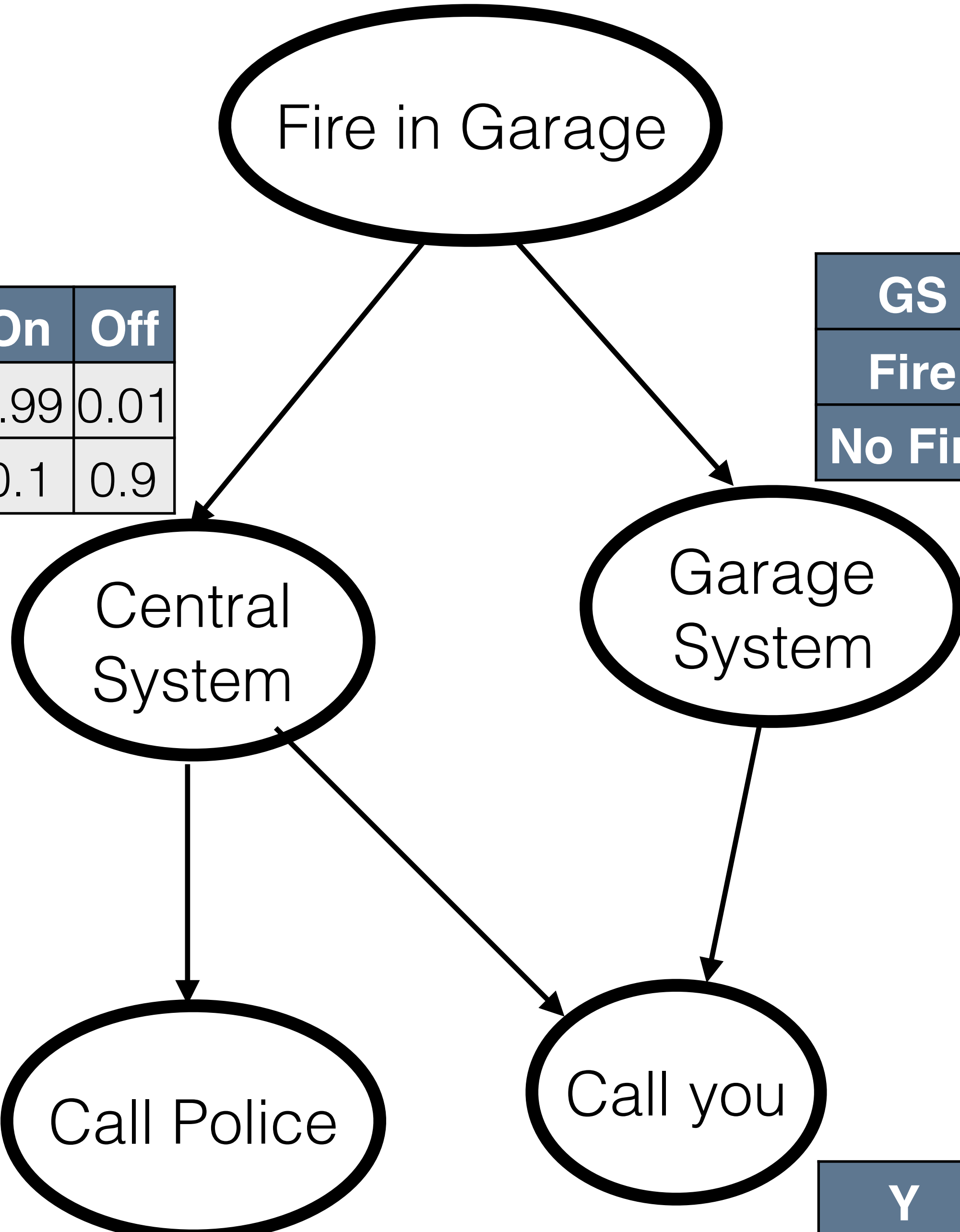
GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

Rejection Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

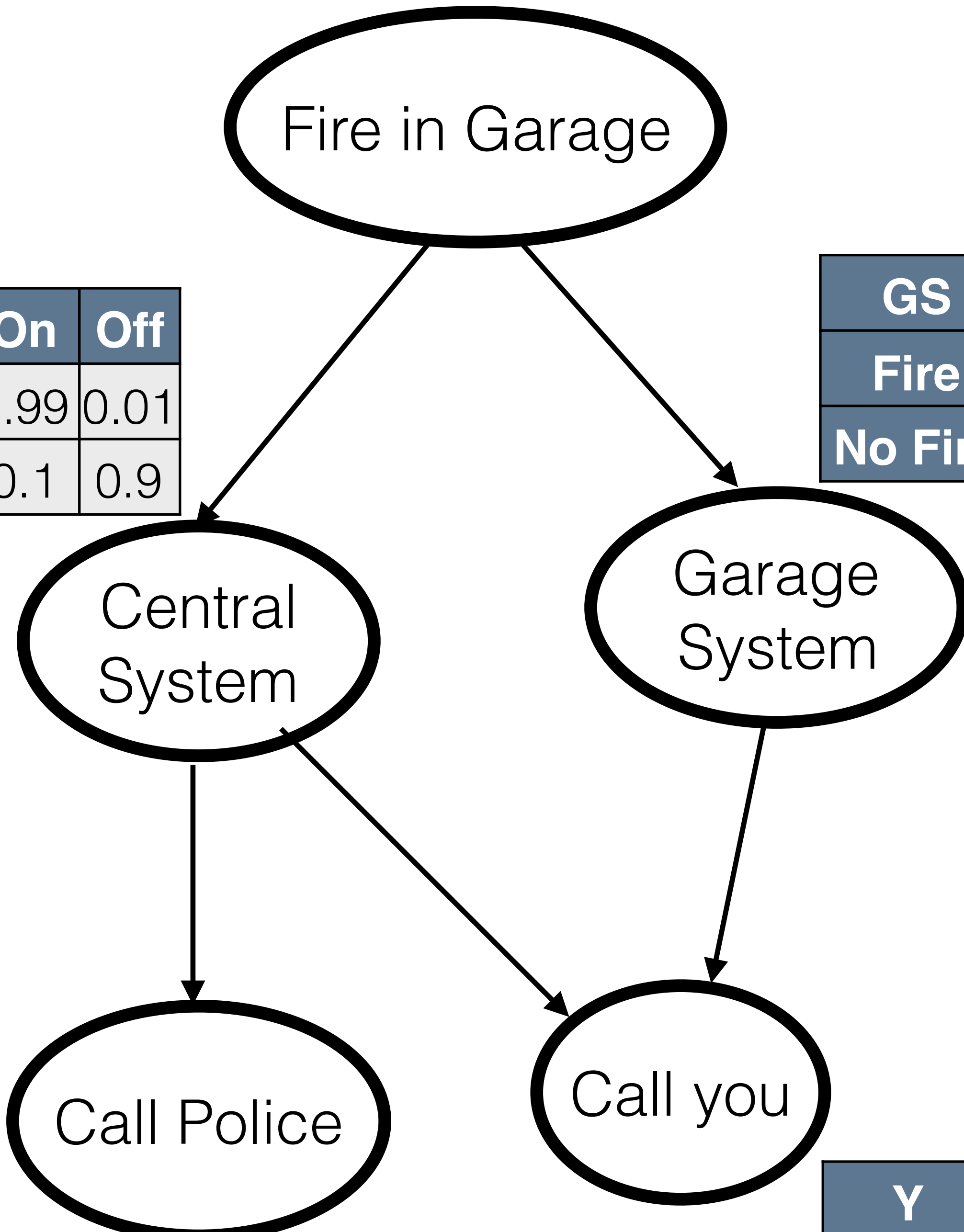
Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

	F	CS	GS	P	Y
1	0				

Rejection Sampling

	F	CS	GS	P	Y
1	0	0			

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

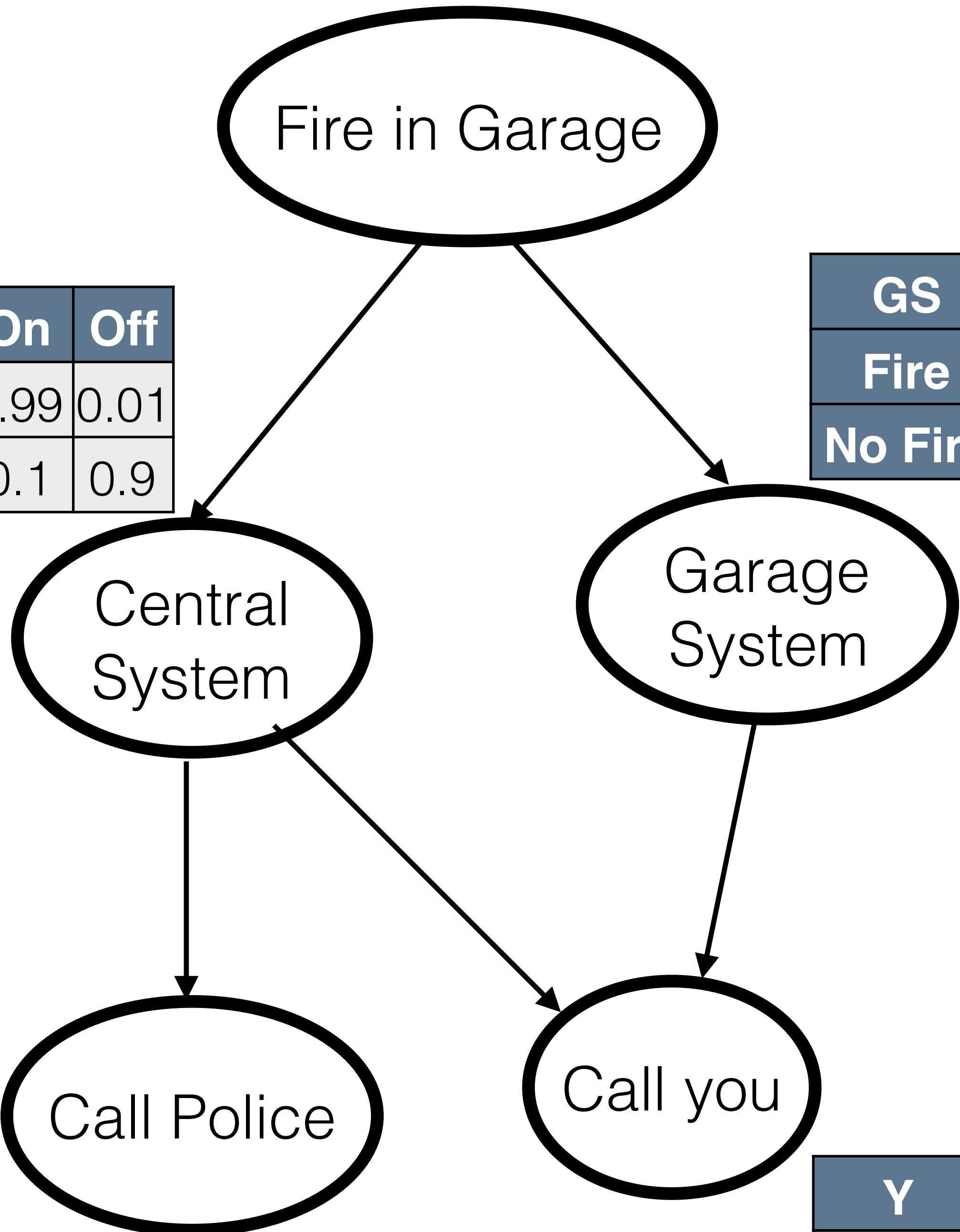
Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Rejection Sampling

	F	CS	GS	P	Y
1	0	0	1		

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

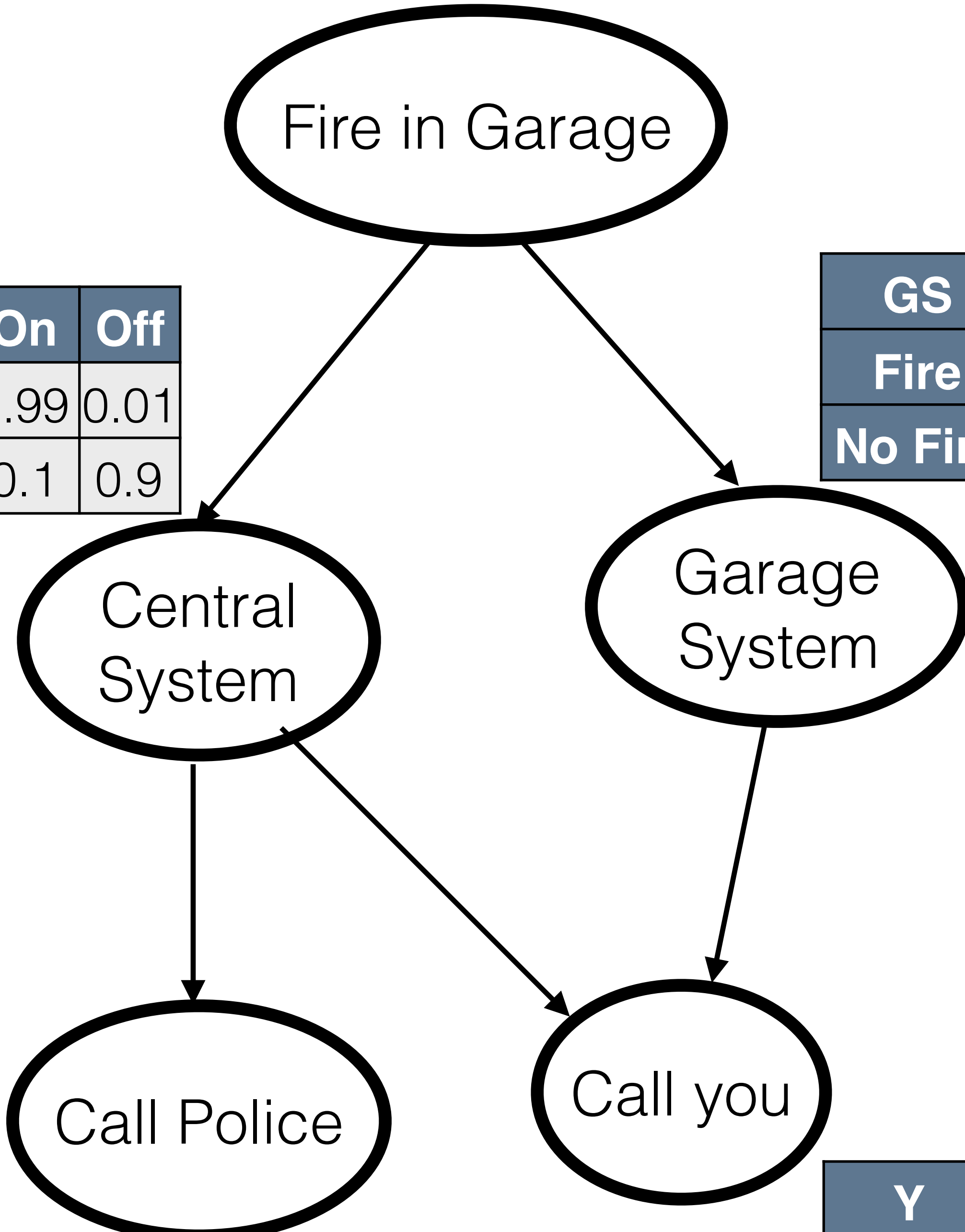
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

REJECTION SAMPLING

Fire	No Fire
0.05	0.95

	F	CS	GS	P	Y
1	0	0	1	0	



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

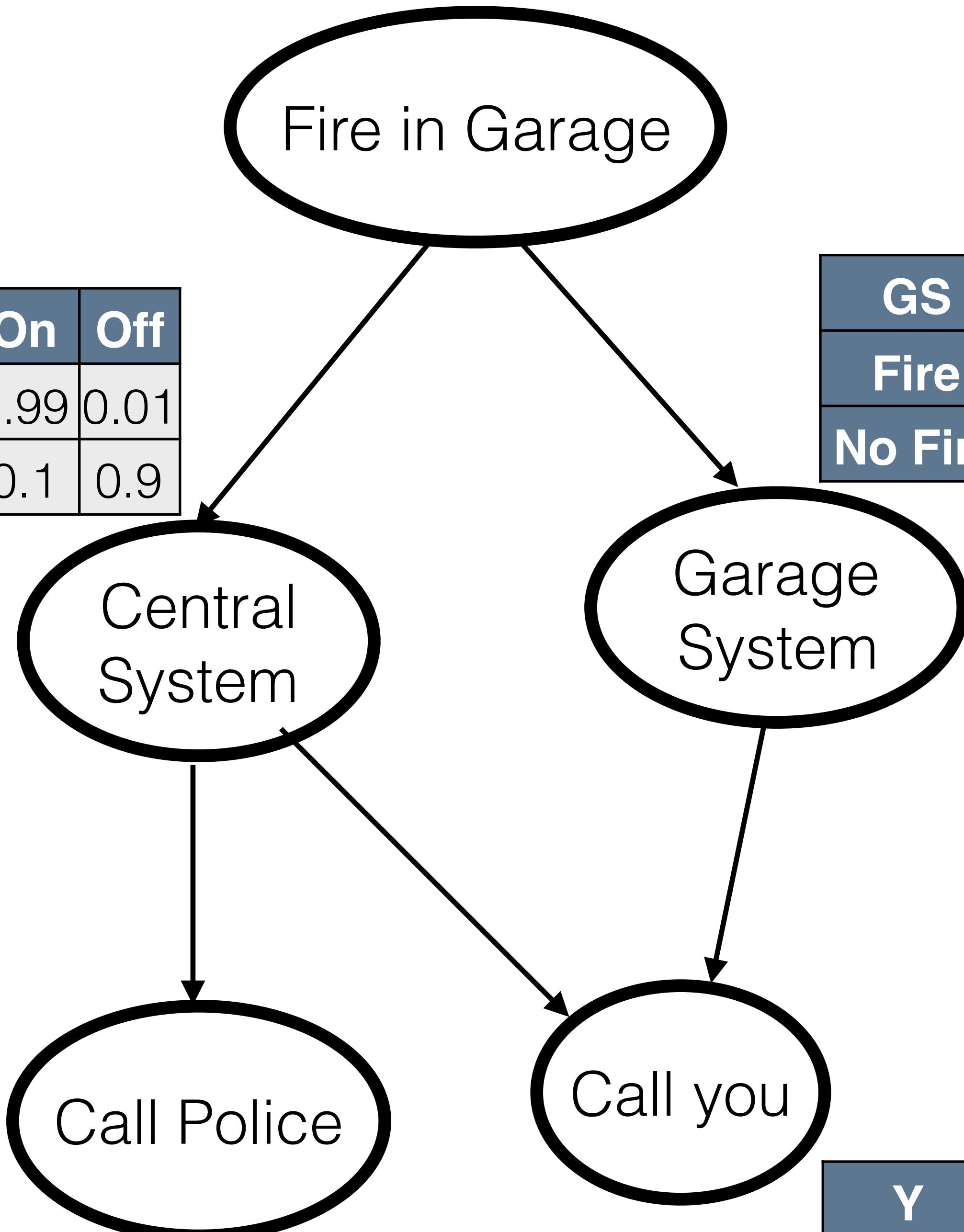
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

Rejection Sampling

Fire	No Fire
0.05	0.95

	F	CS	GS	P	Y
1	0	0	1	0	1



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

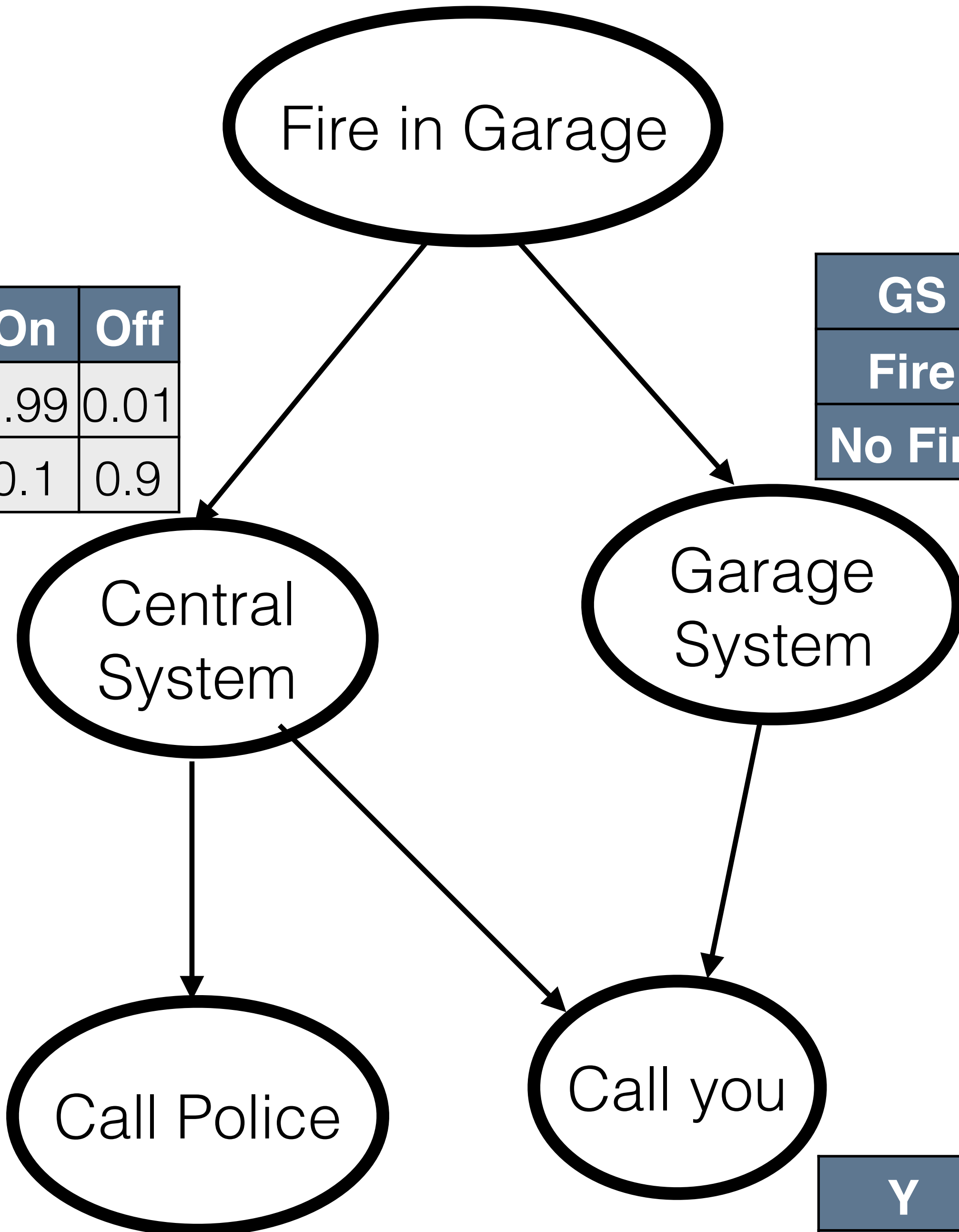
GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Rejection Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

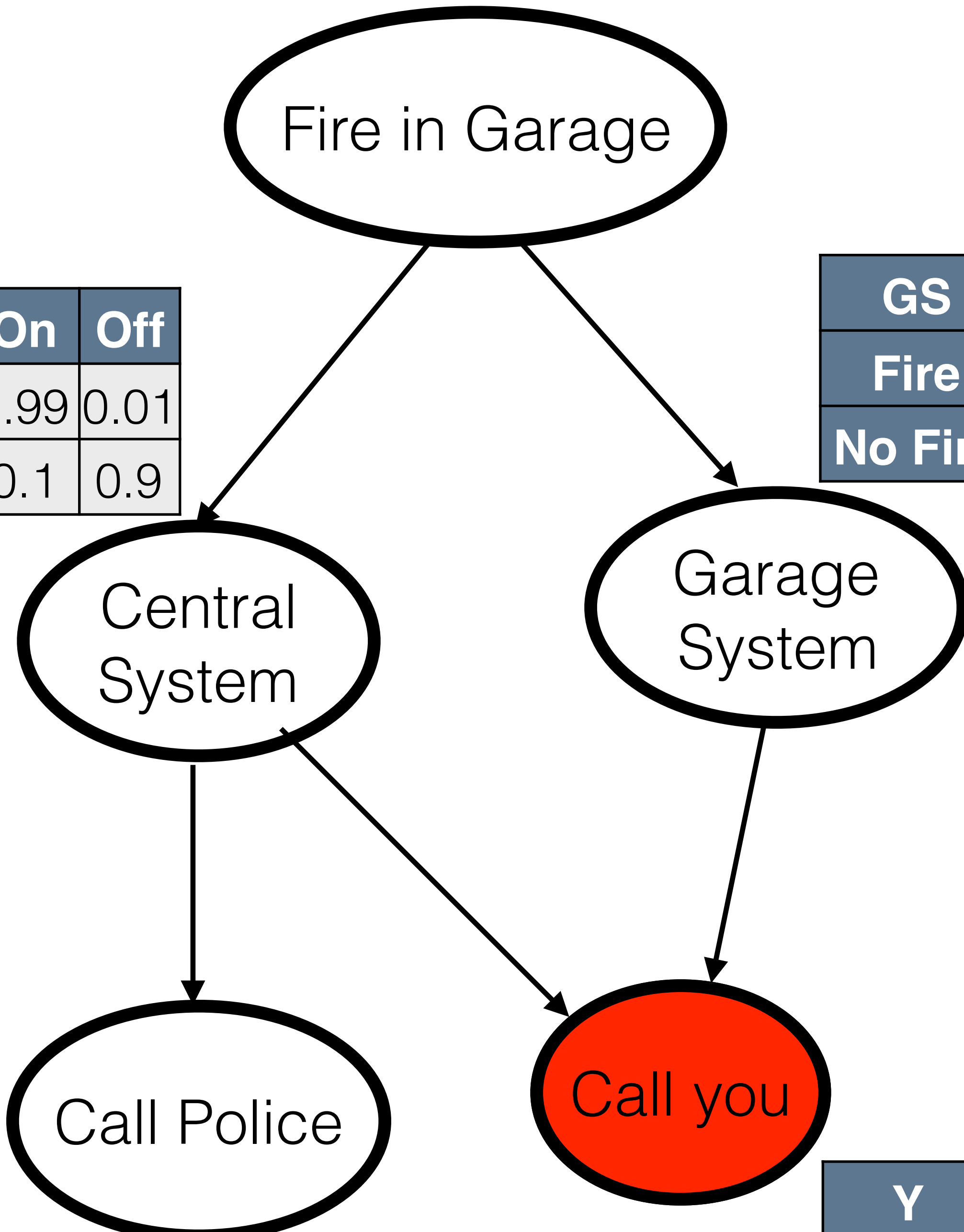
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

REJECTION SAMPLING

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

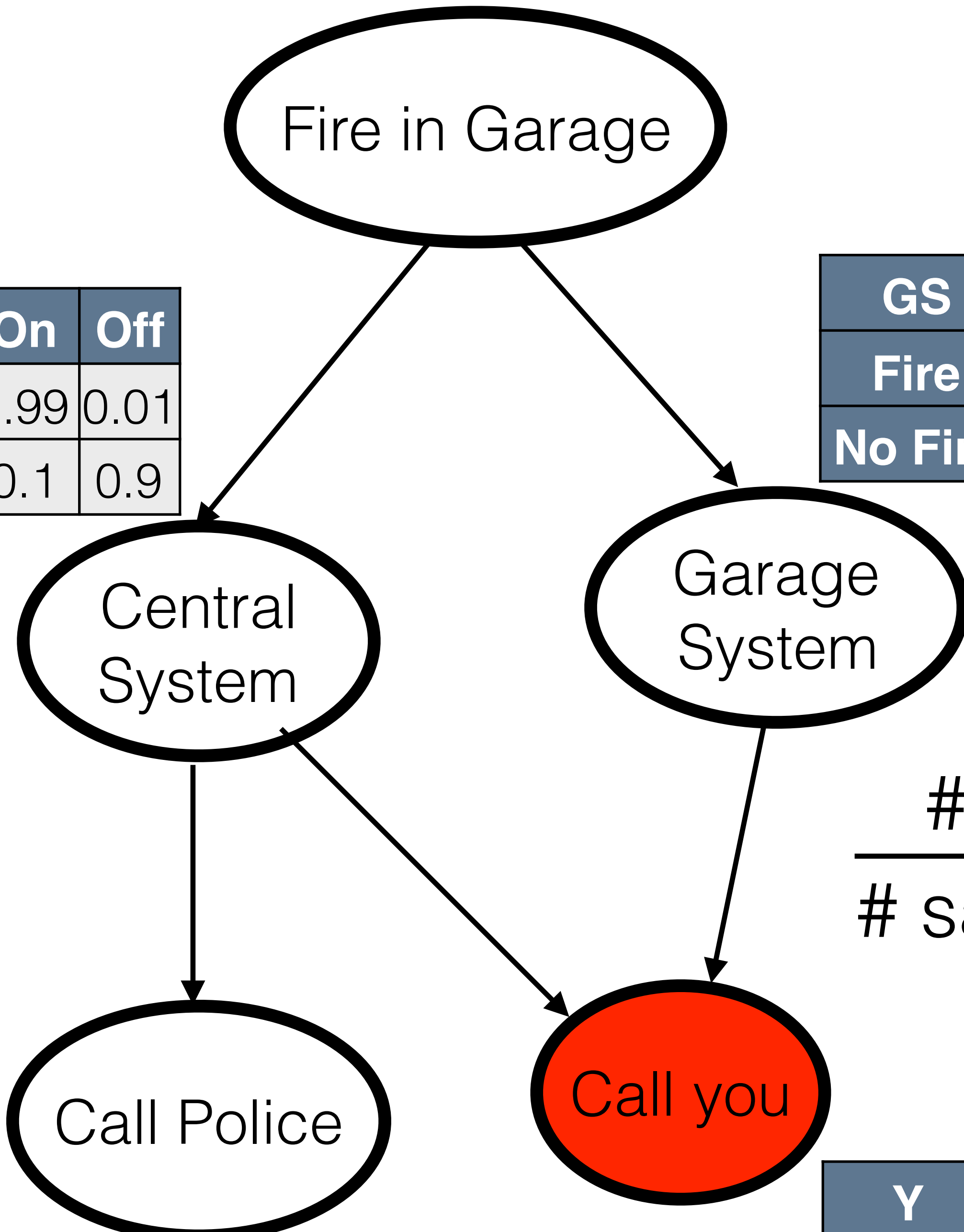
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Rejection Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

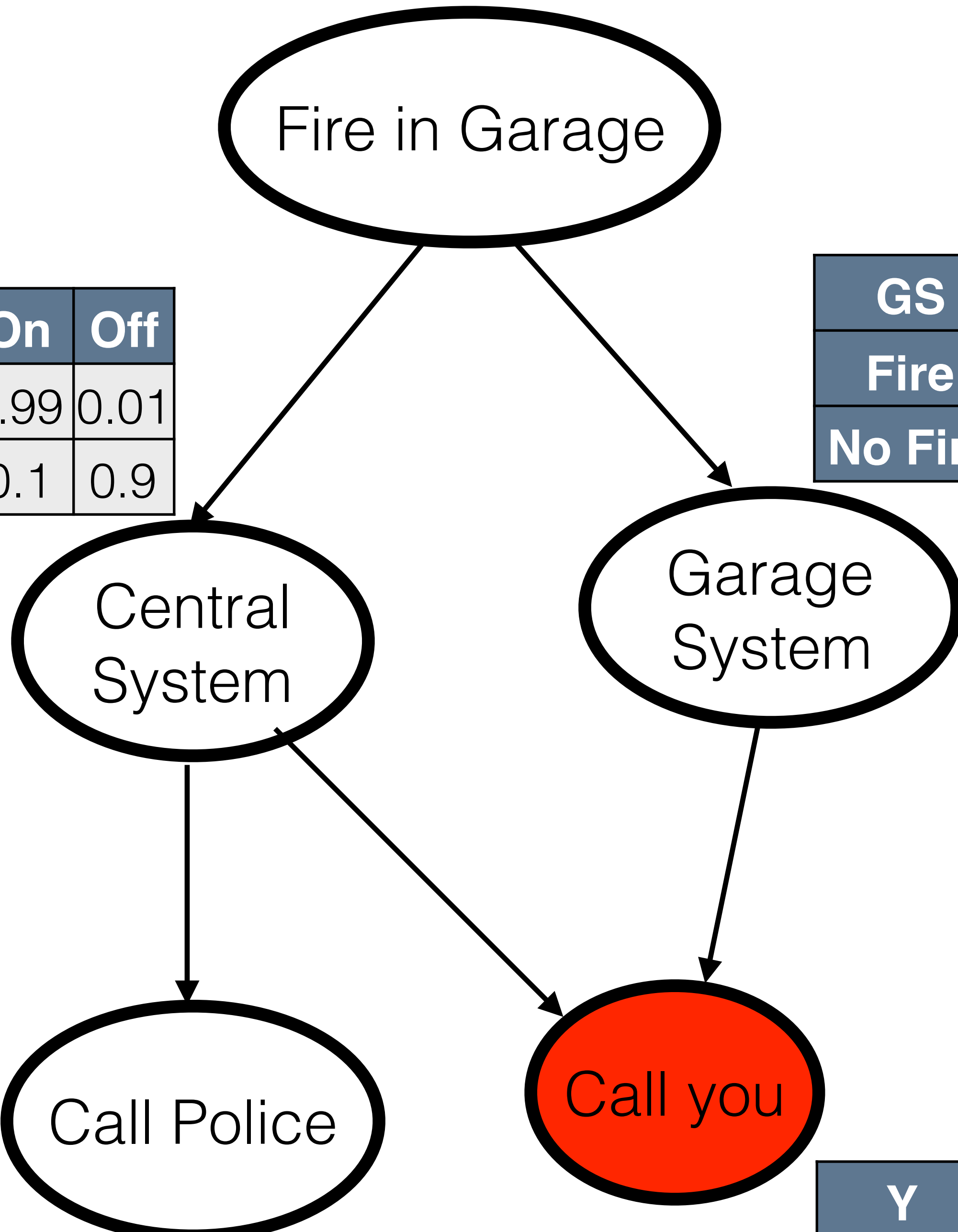
Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

$$\frac{\#\ \text{calls}}{\#\ \text{samples}} = \frac{8}{25}$$

	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Rejection Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

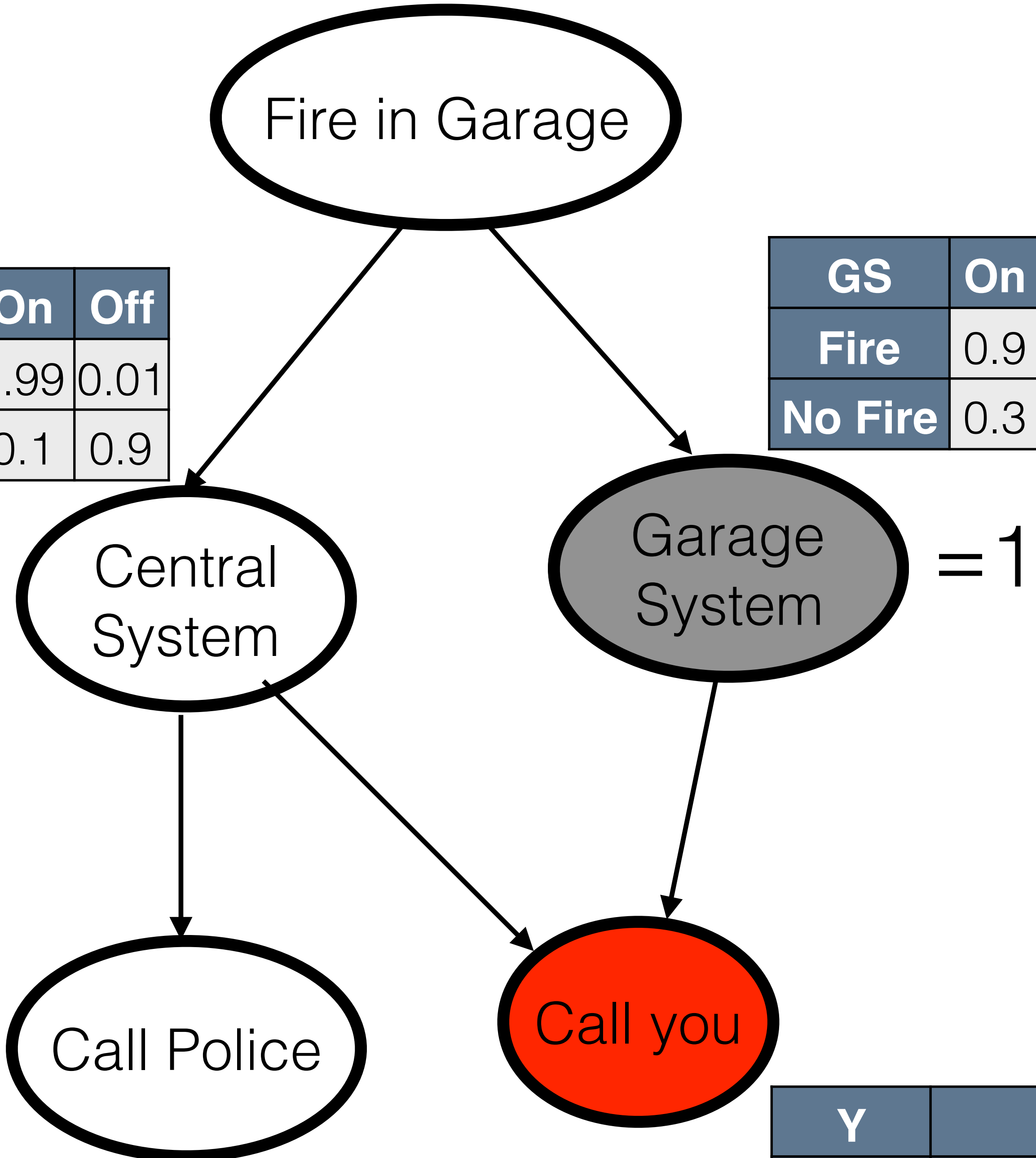
	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Rejection Sampling

Fire	No Fire
0.05	0.95

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7



P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

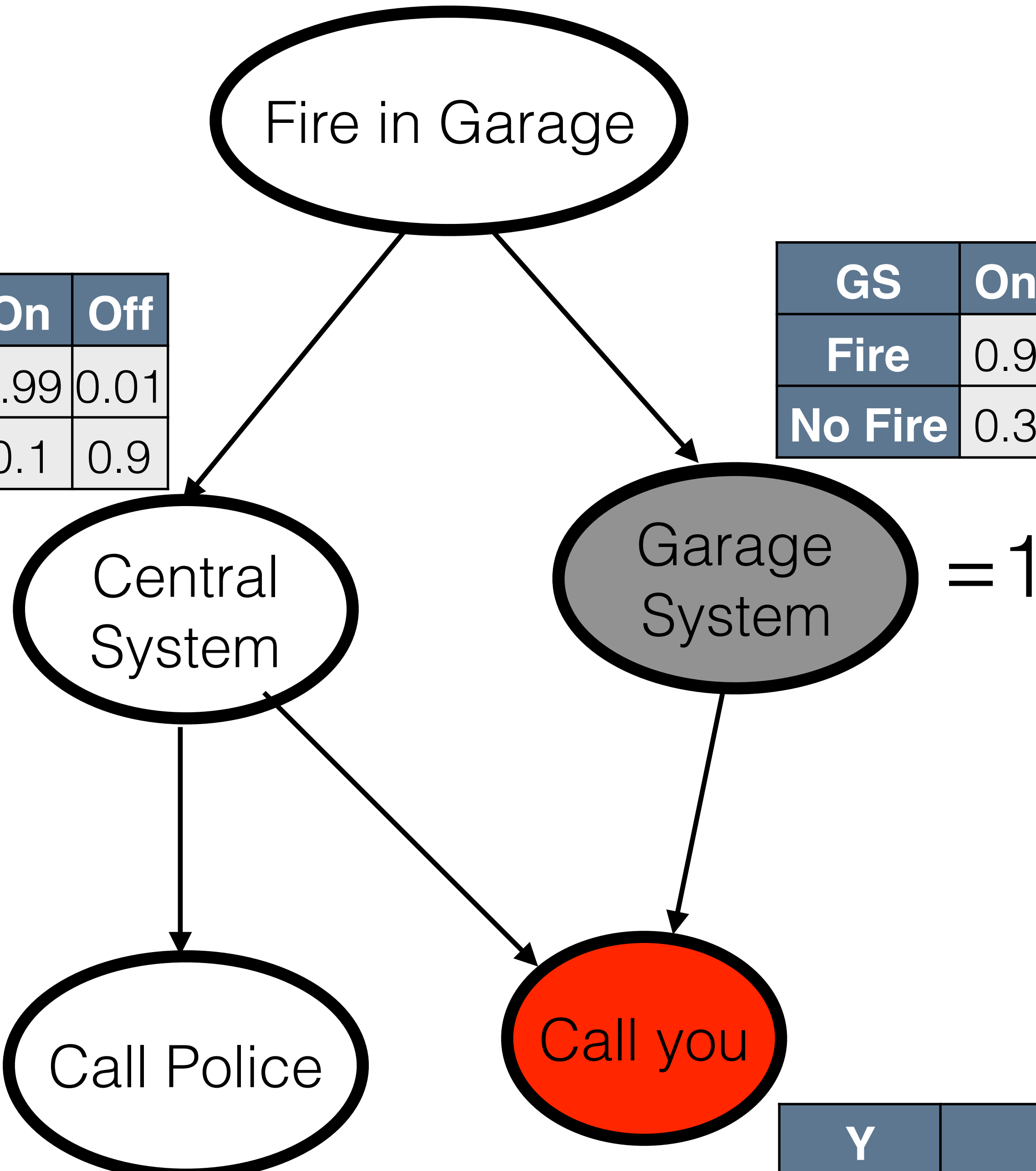
	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Rejection Sampling

Fire	No Fire
0.05	0.95

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7



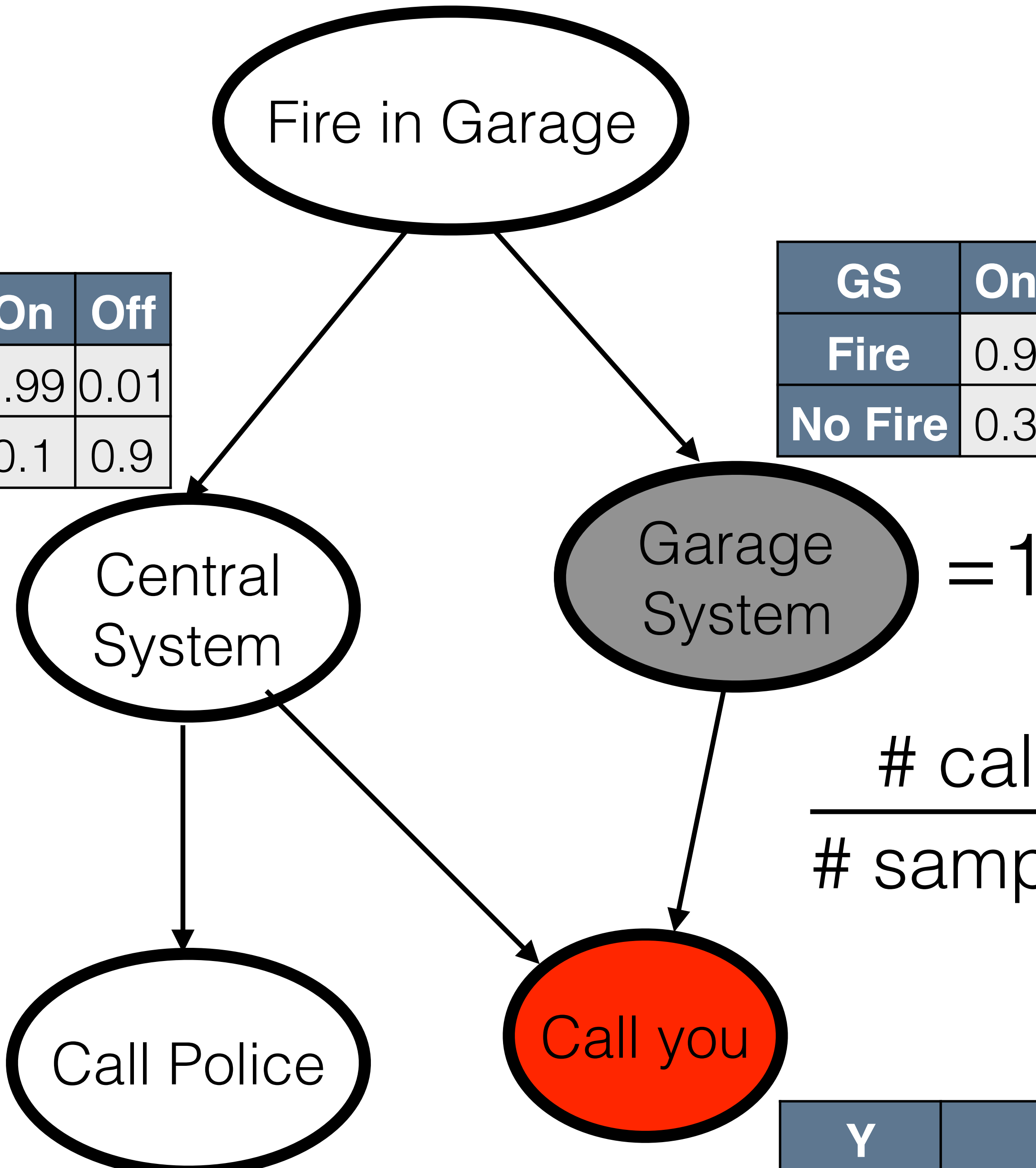
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Rejection Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

$$\frac{\#\text{ calls}}{\#\text{ samples}} = \frac{8}{11}$$

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Rejection Sampling

Algorithm:

Topologically sort variables (parents first children later)

For $t = 1$ to n (number of samples)

For $i = 1$ to N (number of variables in model)

Sample $x_i^t \sim P(X_i|\text{Parents}(X_i) \text{ already sampled})$

End For

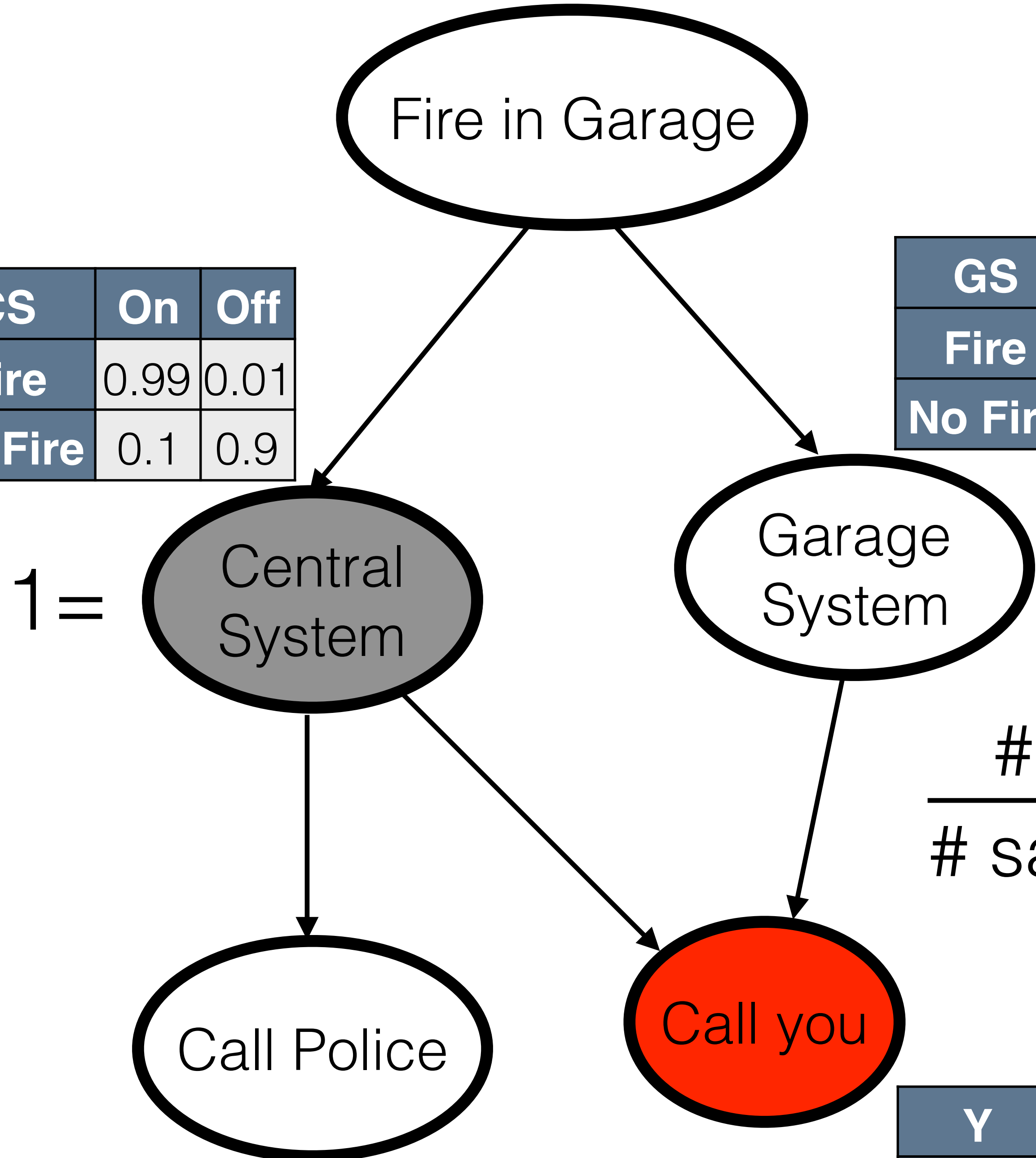
End For

Discard samples that do not match observations

Compute empirical frequencies

Rejection Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

$$\frac{\#\ \text{calls}}{\#\ \text{samples}} = \frac{1}{2}$$

	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

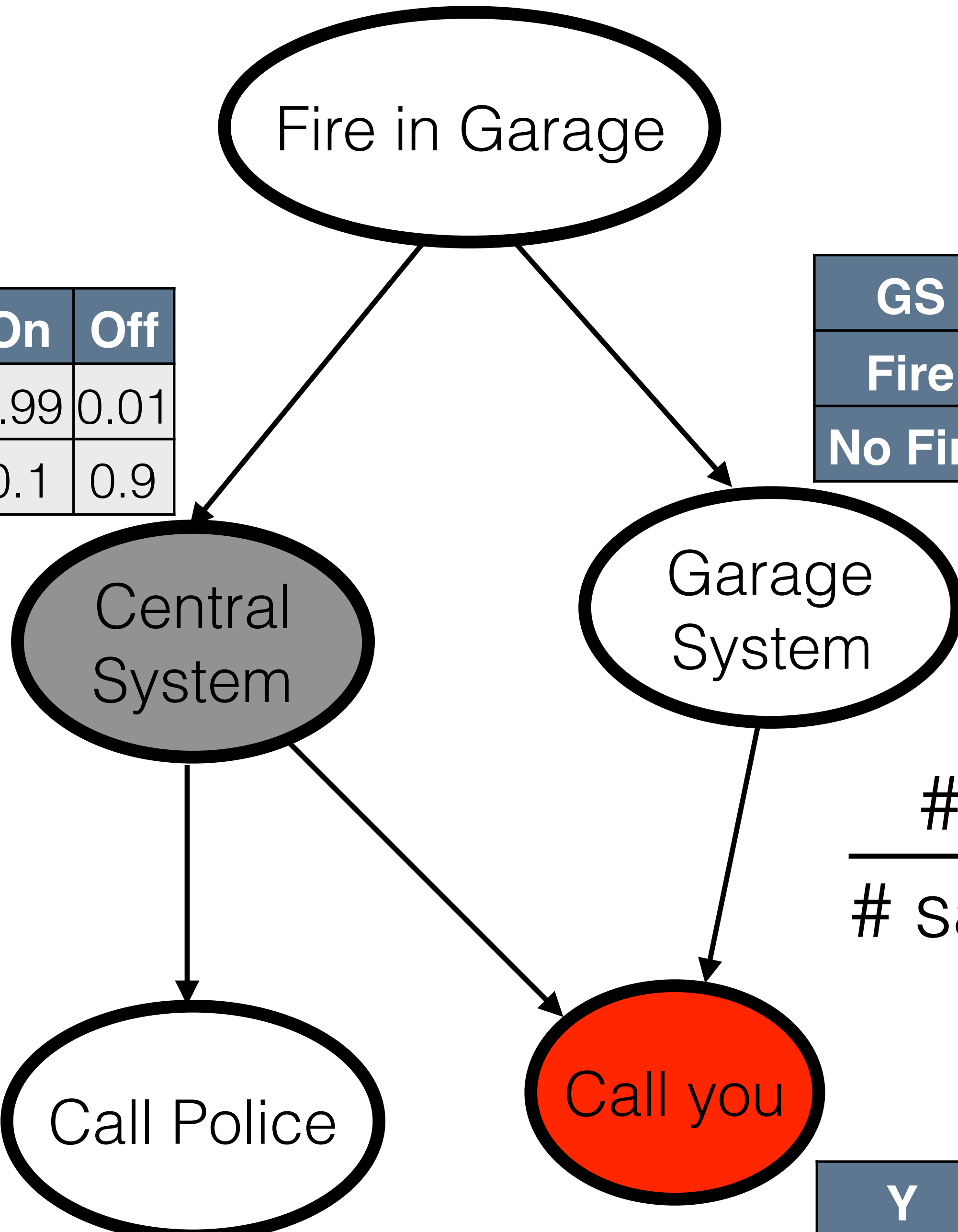
Rejection Sampling

Fire	No Fire
0.05	0.95

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

1=



???!!!

$$\frac{\#\ \text{calls}}{\#\ \text{samples}} = \frac{1}{2}$$

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

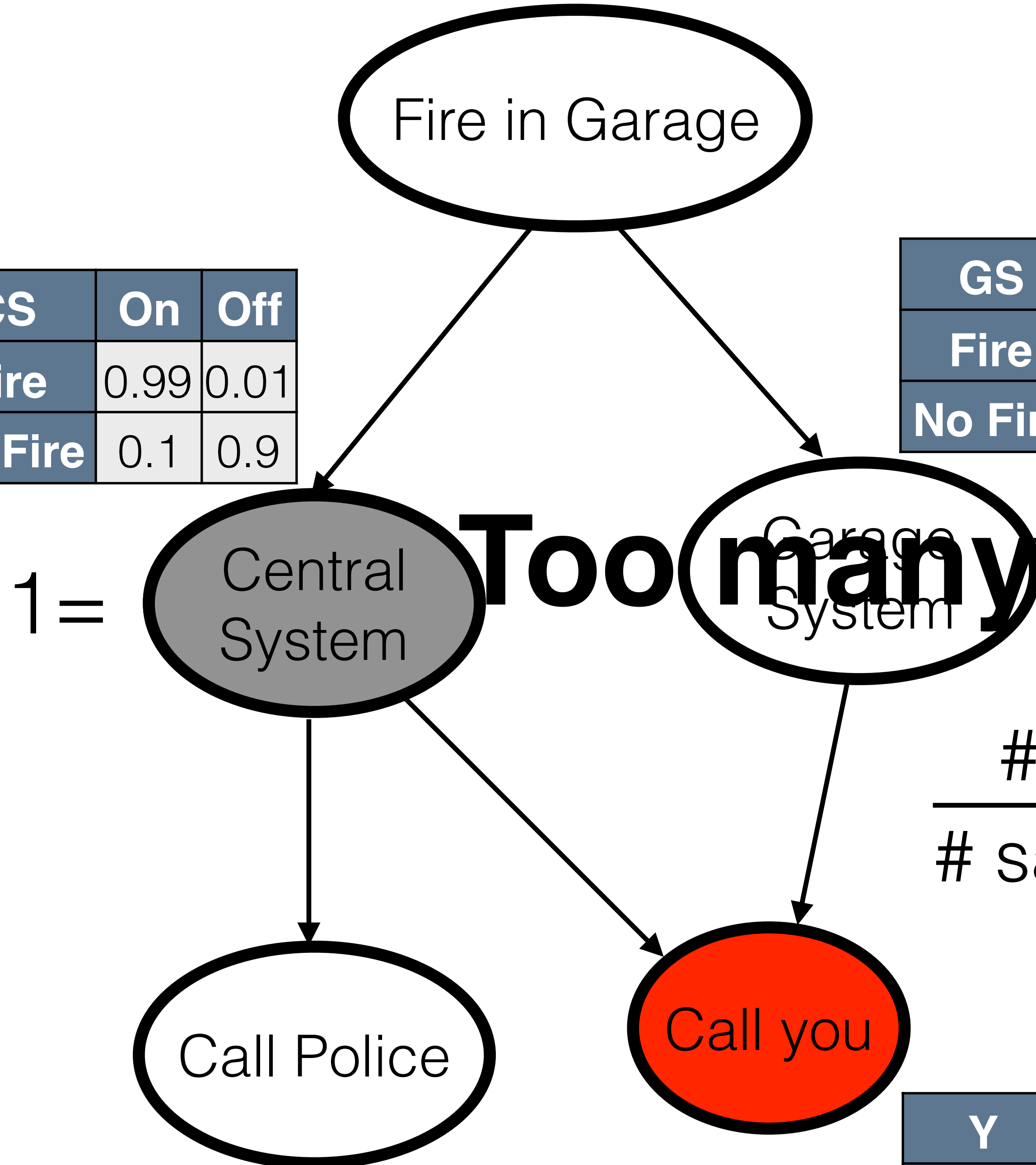
	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	0	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Rejection Sampling

Fire	No Fire
0.05	0.95

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7



Too many rejections :(

?? ? ! ! !

$$\frac{\#\text{ calls}}{\#\text{ samples}} = \frac{1}{2}$$

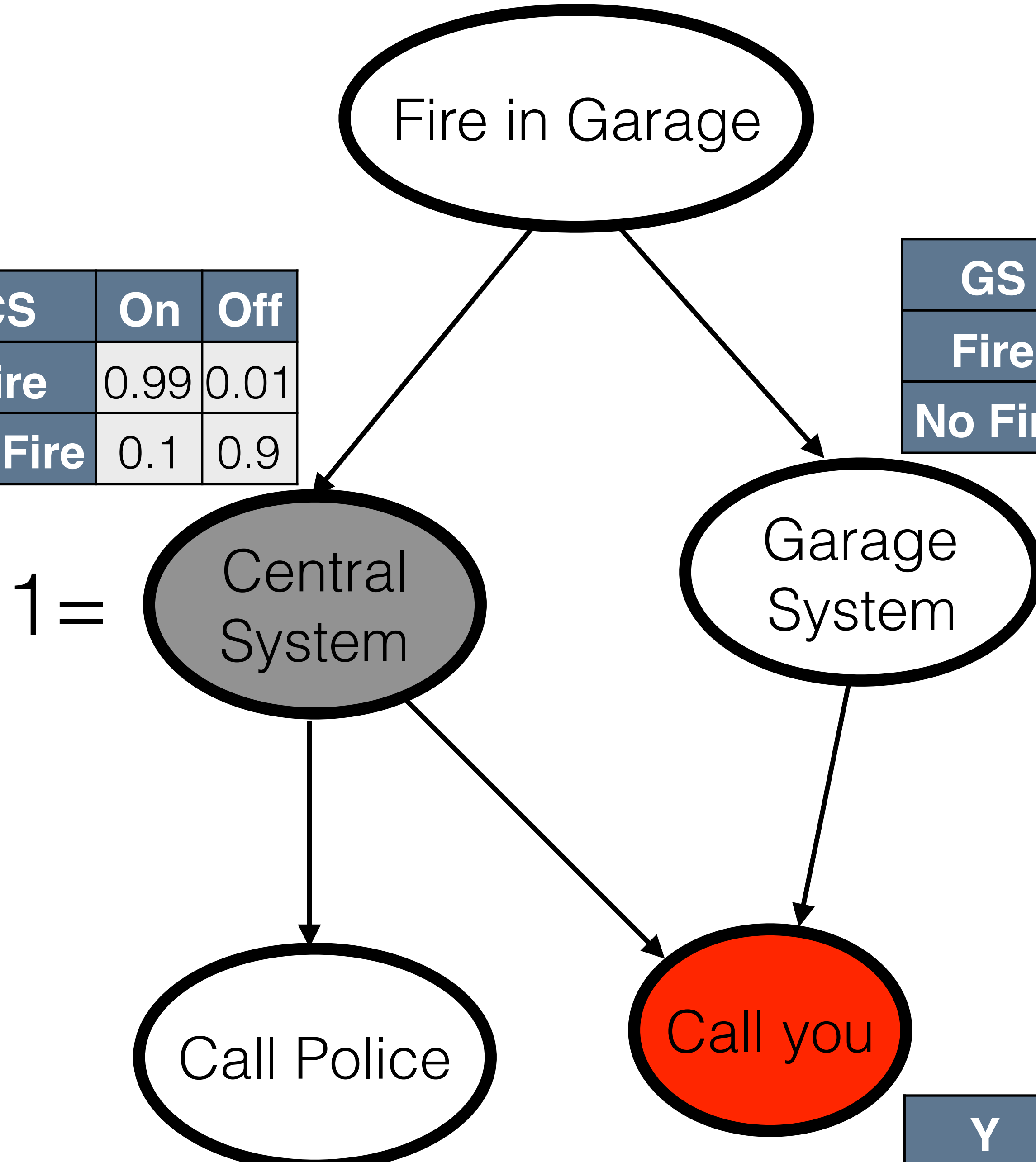
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

	F	CS	GS	P	Y
1	0	0	1	0	1
2	0	1	0	1	0
3	1	1	1	1	1
4	0	0	0	0	0
5	0	0	1	0	0
6	0	0	1	0	1
7	0	0	0	0	0
8	0	0	1	0	1
9	0	0	0	0	0
10	0	0	1	0	1
11	[illegible]	0	1	0	0
12	0	0	1	0	1
13	0	0	1	0	1
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	1	0	0
17	0	0	0	0	0
18	0	0	0	0	0
19	0	0	0	0	0
20	0	0	0	0	0
21	0	0	0	0	0
22	0	0	0	0	0
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	1	0	1

Importance Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

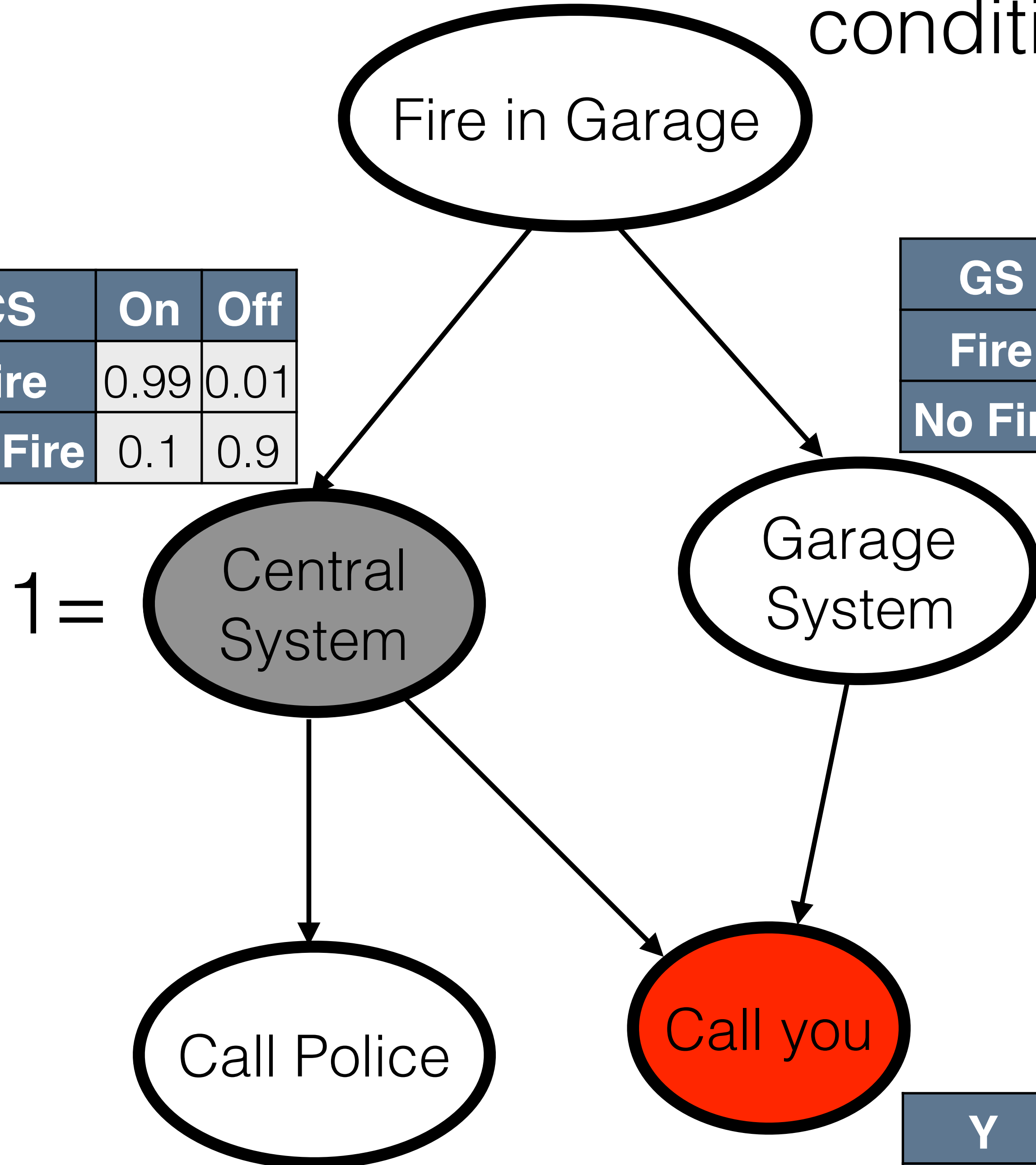
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

Importance Sampling

If we could sample directly from distribution conditioned on CS = 1 it would be great!!!

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

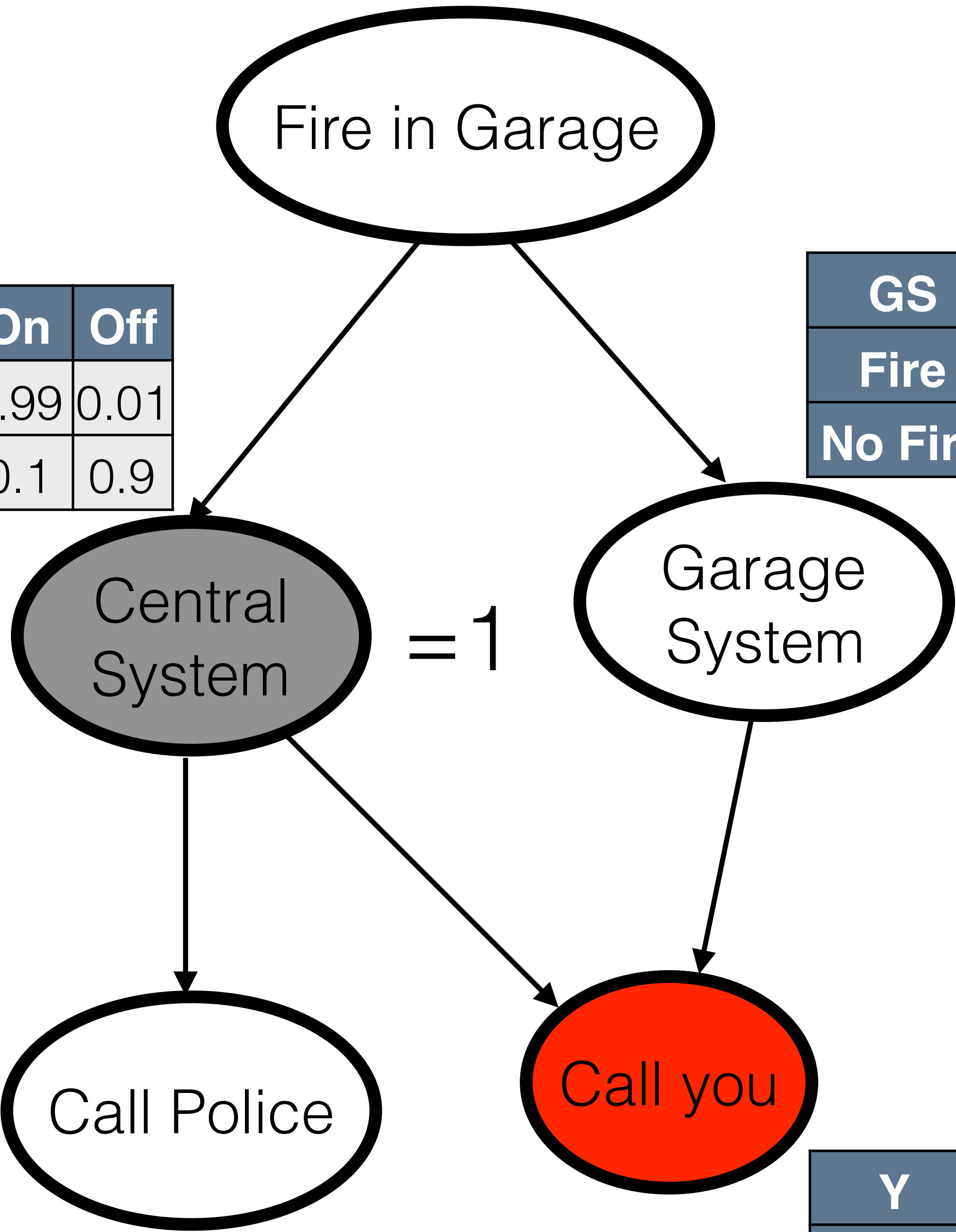
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

Importance Sampling

Fire	No Fire
0.05	0.95

	F	CS	GS	P	Y
1					



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

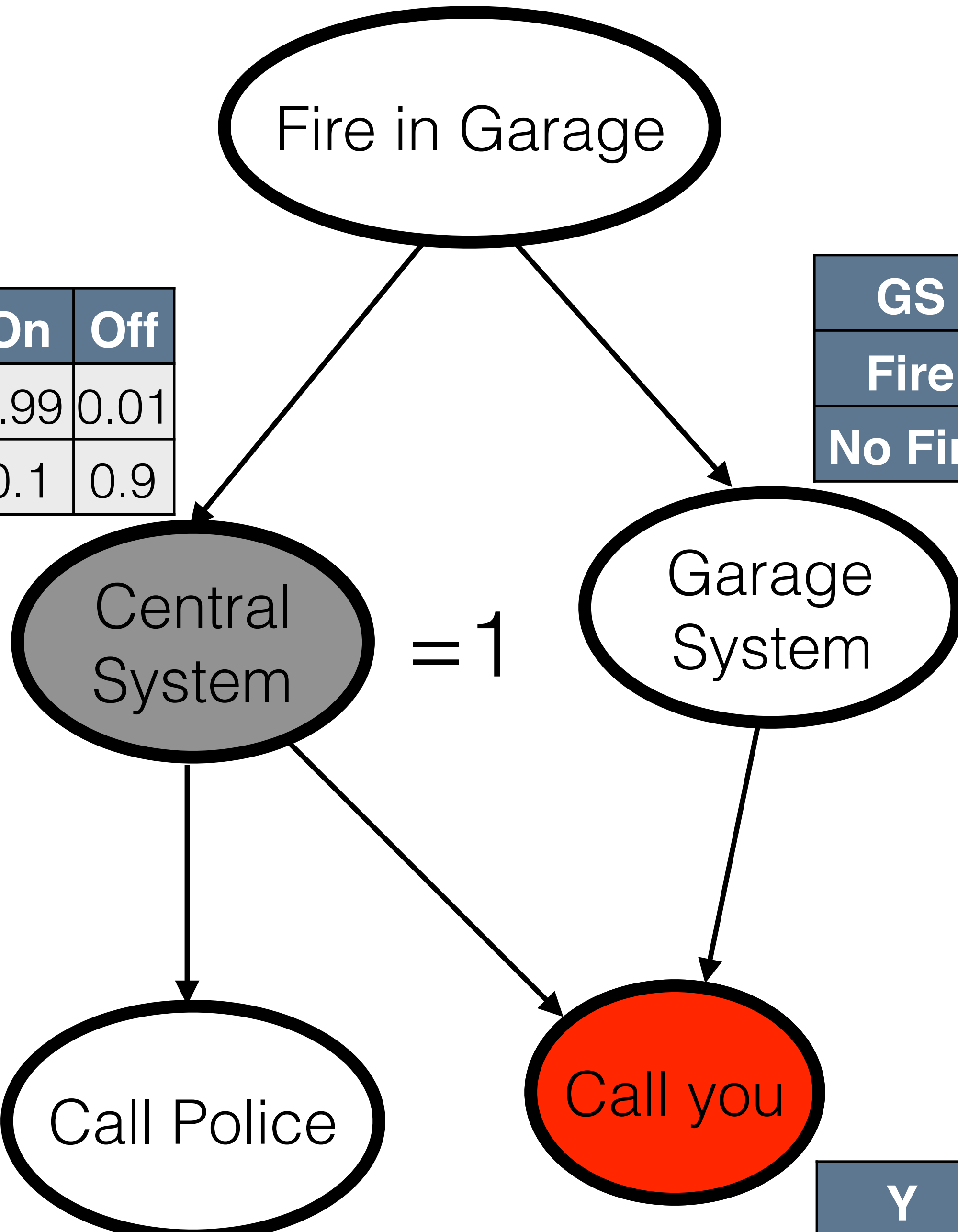
GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

IMPORTANCE SAMPLING

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

	F	CS	GS	P	Y
1	0				

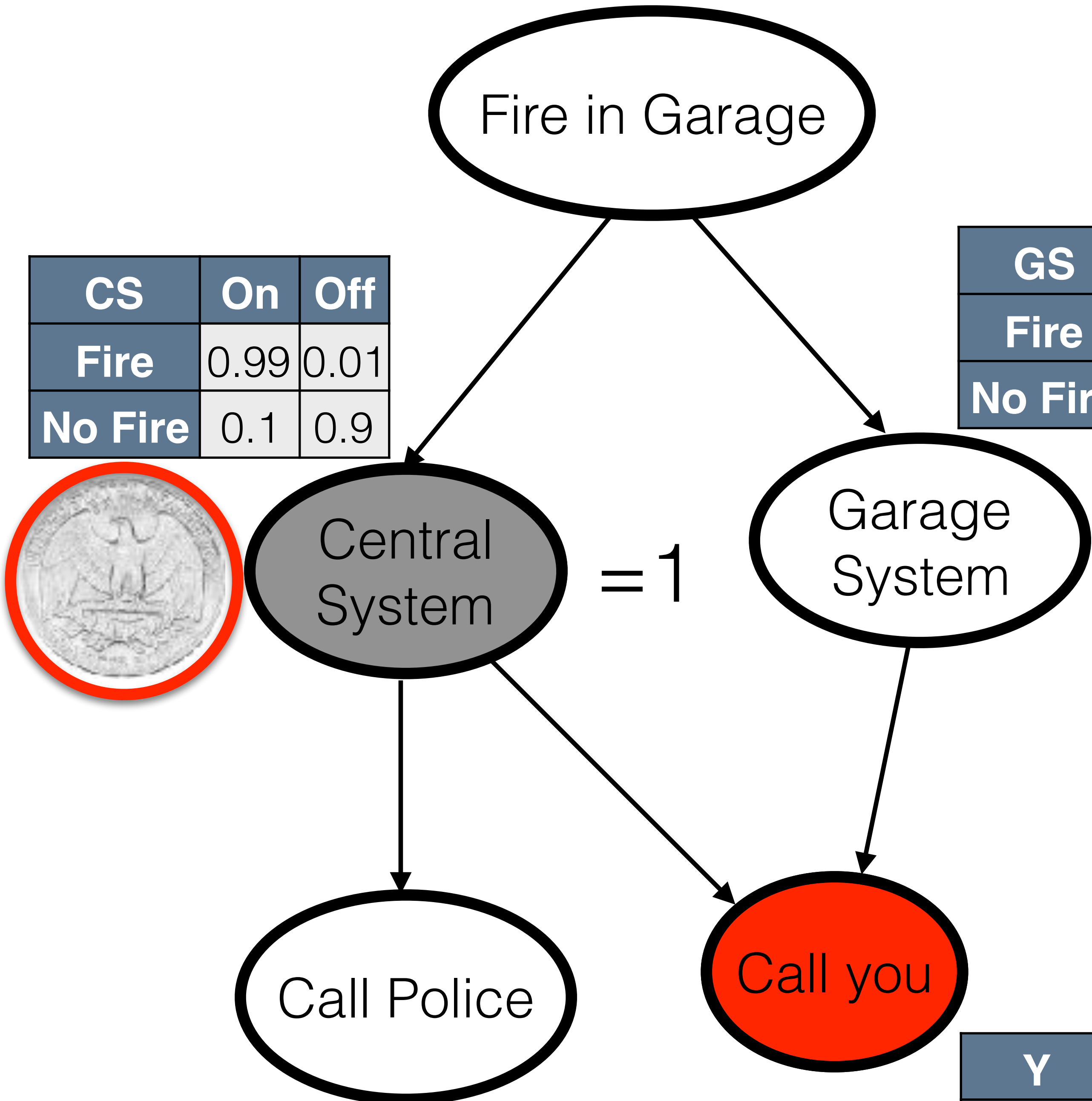
IMPORTANCE SAMPLING

	F	CS	GS	P	Y
1	0	1			

Fire	No Fire
0.05	0.95

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7



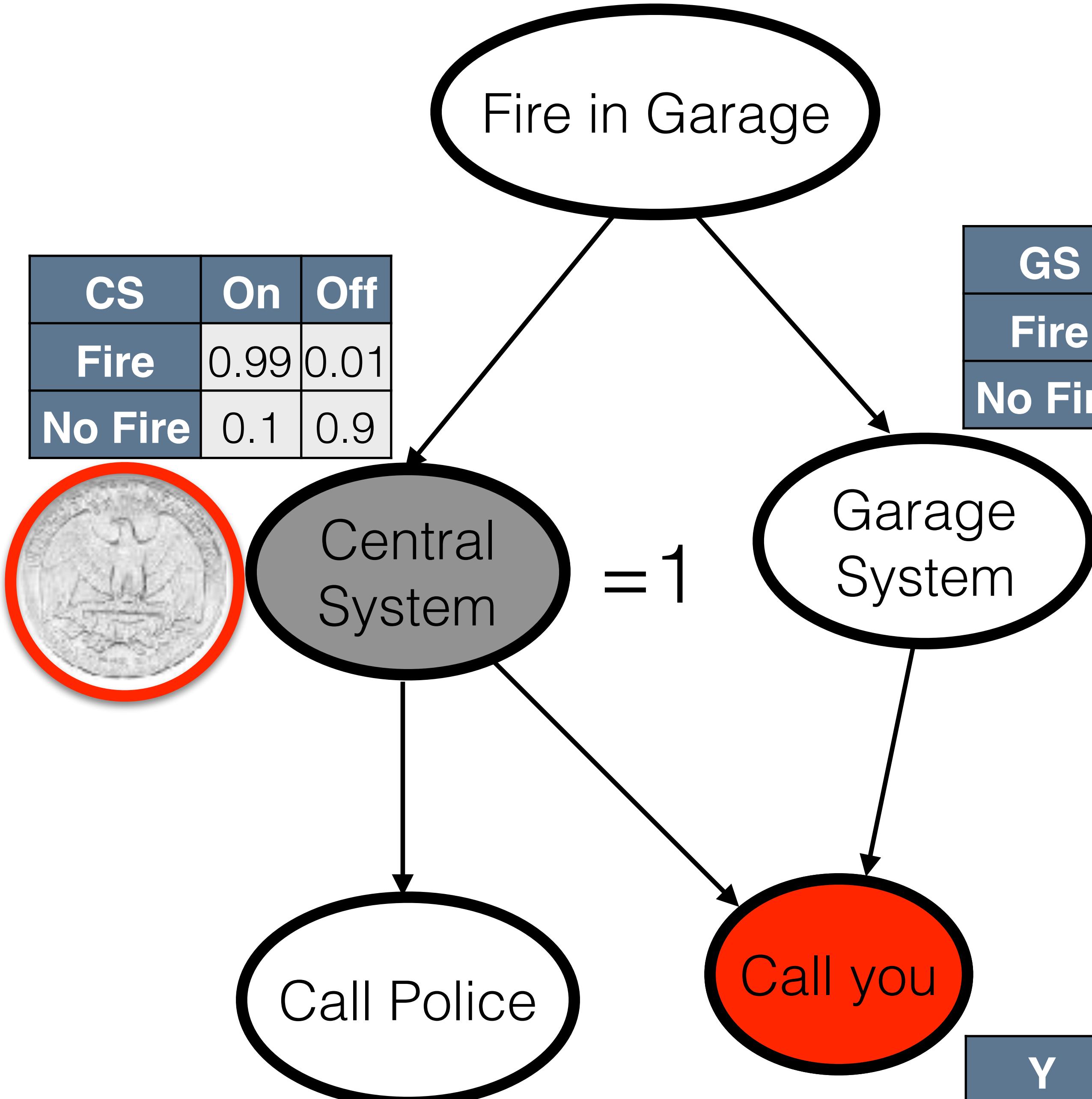
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

Importance Sampling

Fire	No Fire
0.05	0.95

	F	CS	GS	P	Y
1	0	1	0		



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

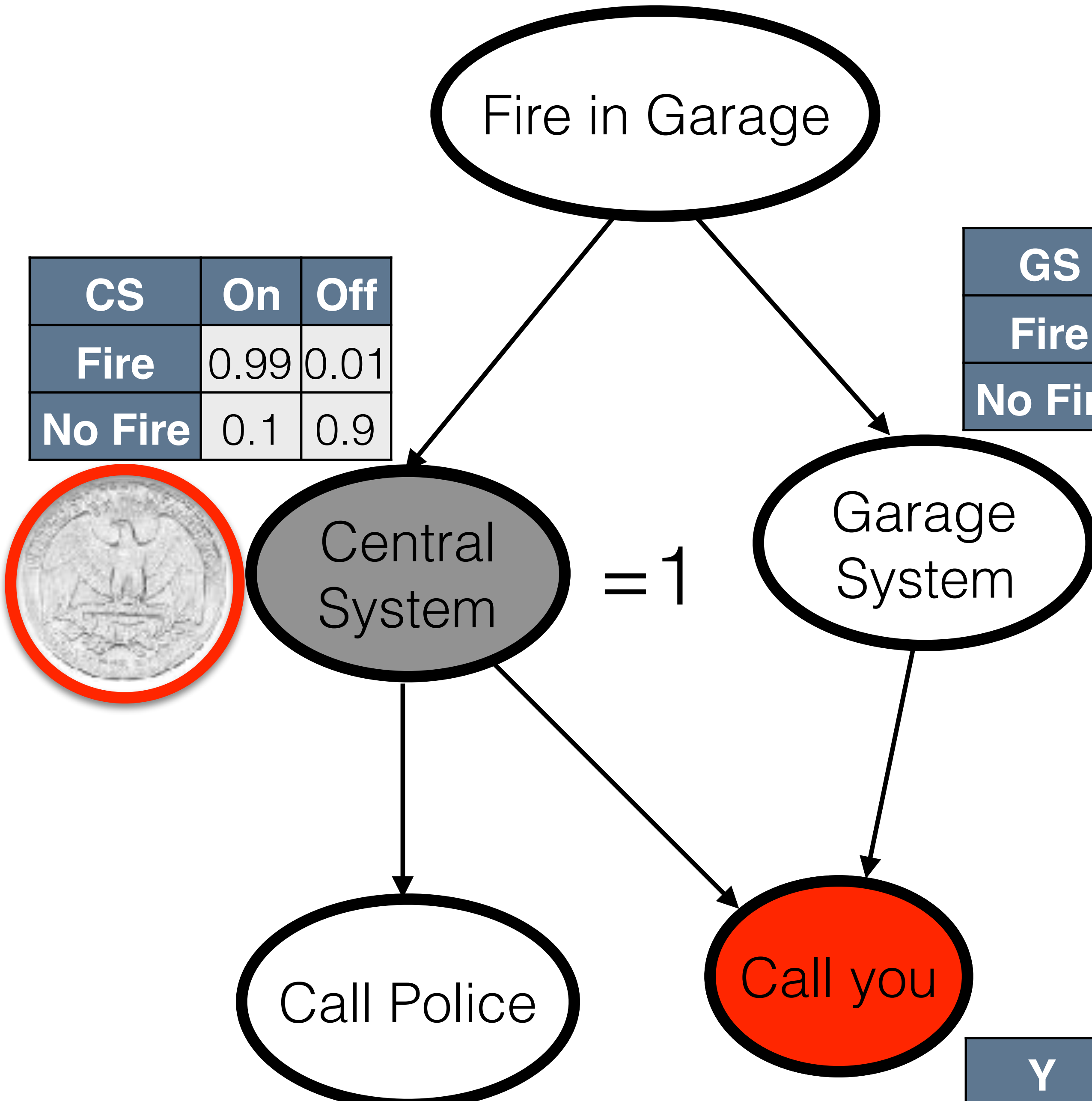
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

Importance Sampling

	F	CS	GS	P	Y
1	0	1	0	1	

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

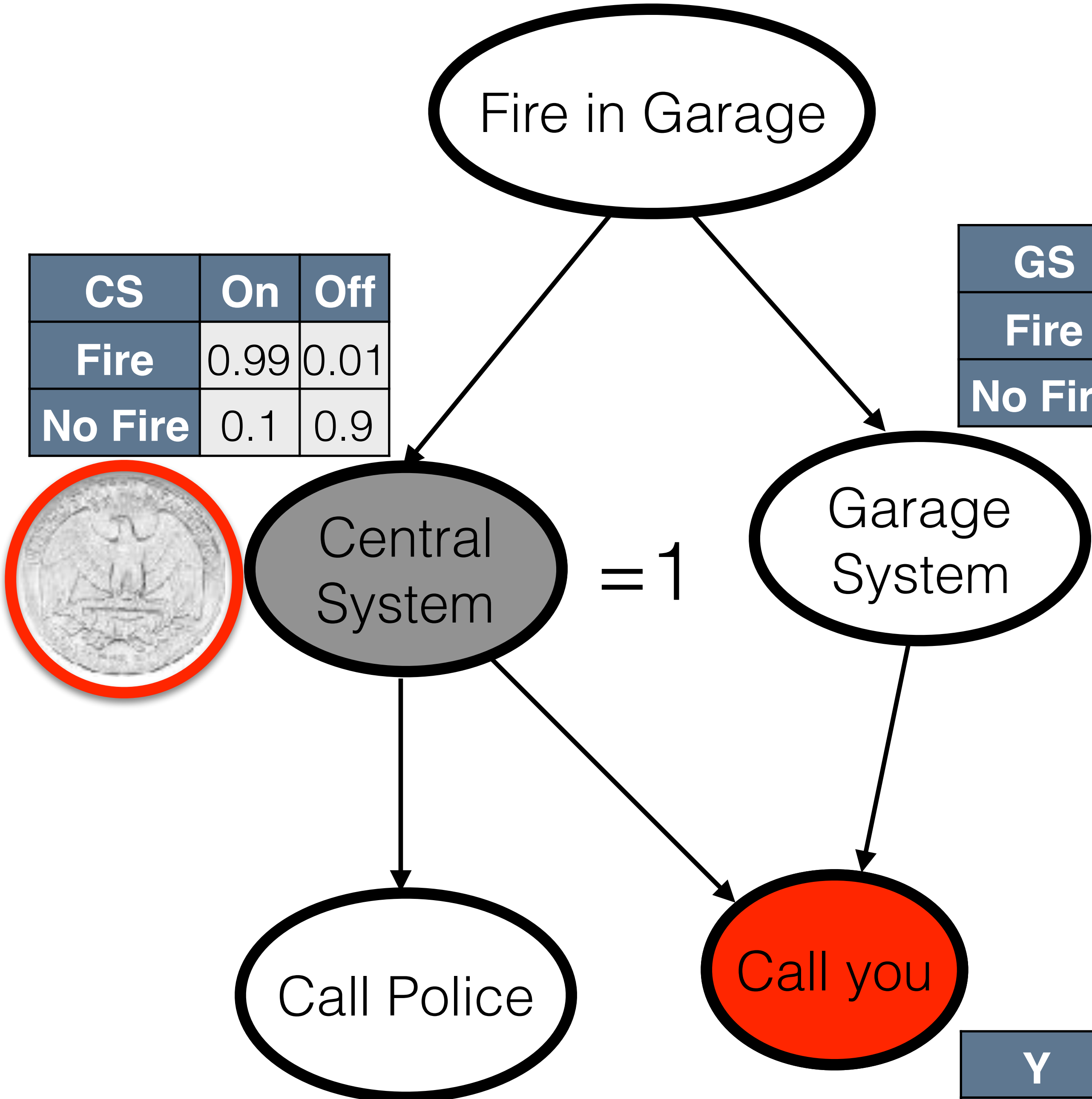
Importance Sampling

Fire	No Fire
0.05	0.95

	F	CS	GS	P	Y
1	0	1	0	1	0

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7



P	Call	no call
CS on	1	0
CS off	0.01	0.99

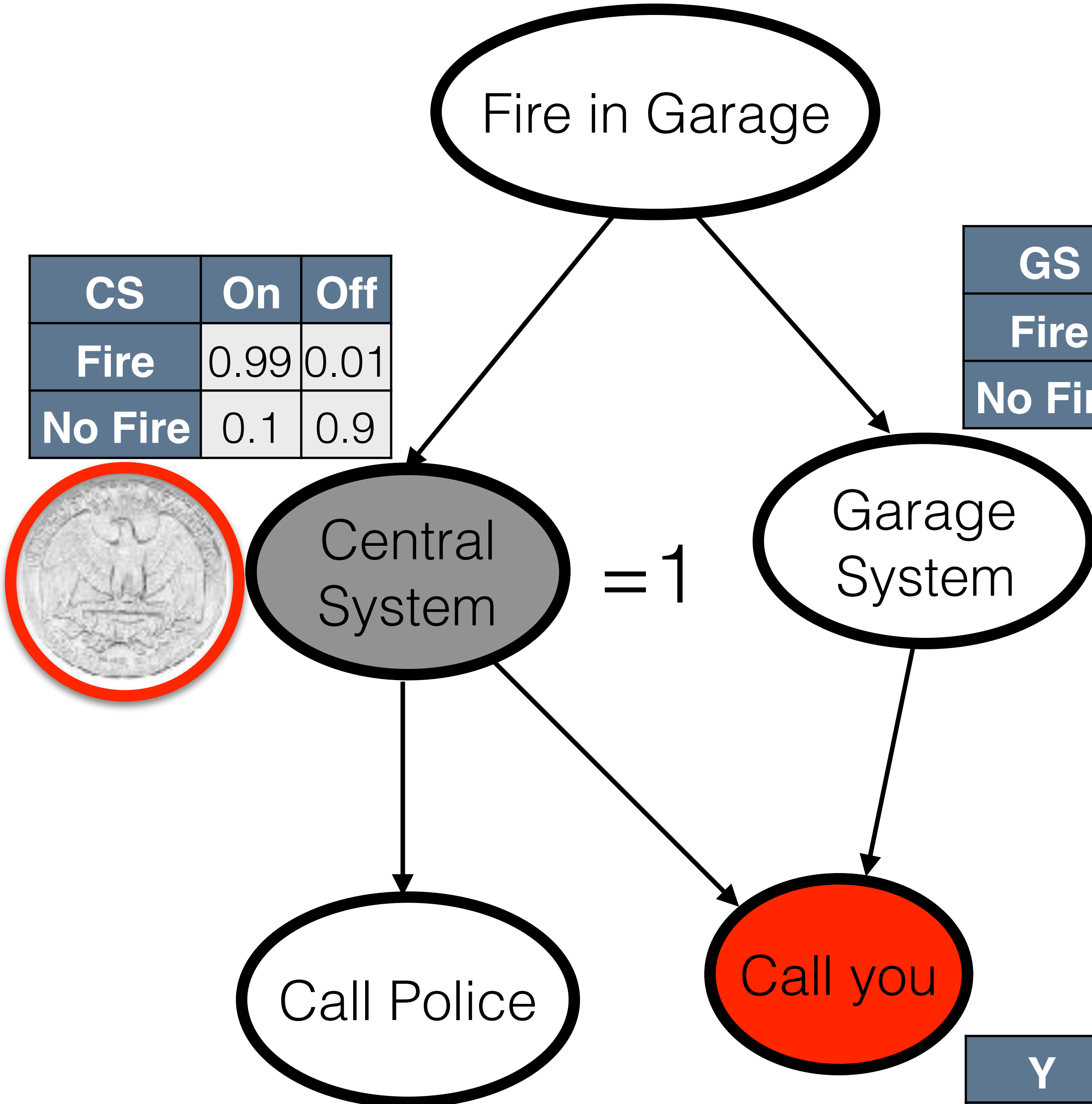
Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

IMPORTANCE SAMPLING

Fire	No Fire
0.05	0.95

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7



P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no call
CS on	GS on	0.95	0.05
CS off	GS on	0.7	0.3
CS on	GS off	0.9	0.1
CS off	GS off	0.01	0.99

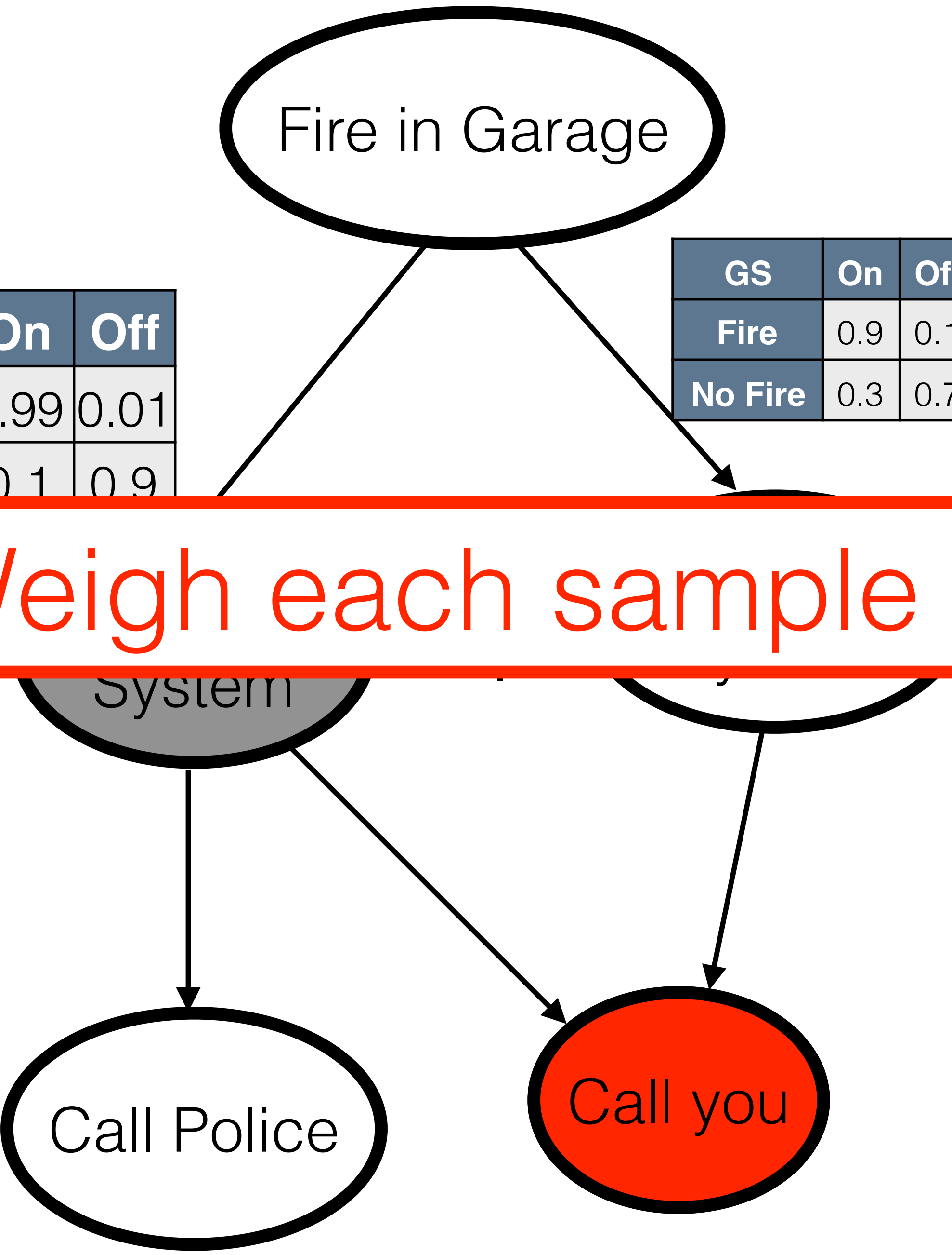
	F	CS	GS	P	Y
1	0	1	0	1	0
2	0	1	1	1	1
3	1	1	1	1	1
4	0	1	0	1	1
5	0	1	0	1	1
6	0	1	0	1	1
7	0	1	1	1	1
8	0	1	0	1	0
9	0	1	0	1	1
10	0	1	1	1	1
11	0	1	0	1	1
12	0	1	1	1	0
13	0	1	0	1	1
14	0	1	1	1	1
15	0	1	0	1	1
16	0	1	0	1	1
17	0	1	0	1	1
18	0	1	0	1	1
19	0	1	1	1	1
20	0	1	0	1	1
21	0	1	0	1	1
22	0	1	0	1	1
23	0	1	1	1	1
24	0	1	0	1	1
25	0	1	1	1	1

Importance Sampling

Fire	No Fire
0.05	0.95

CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7



Weigh each sample differently!

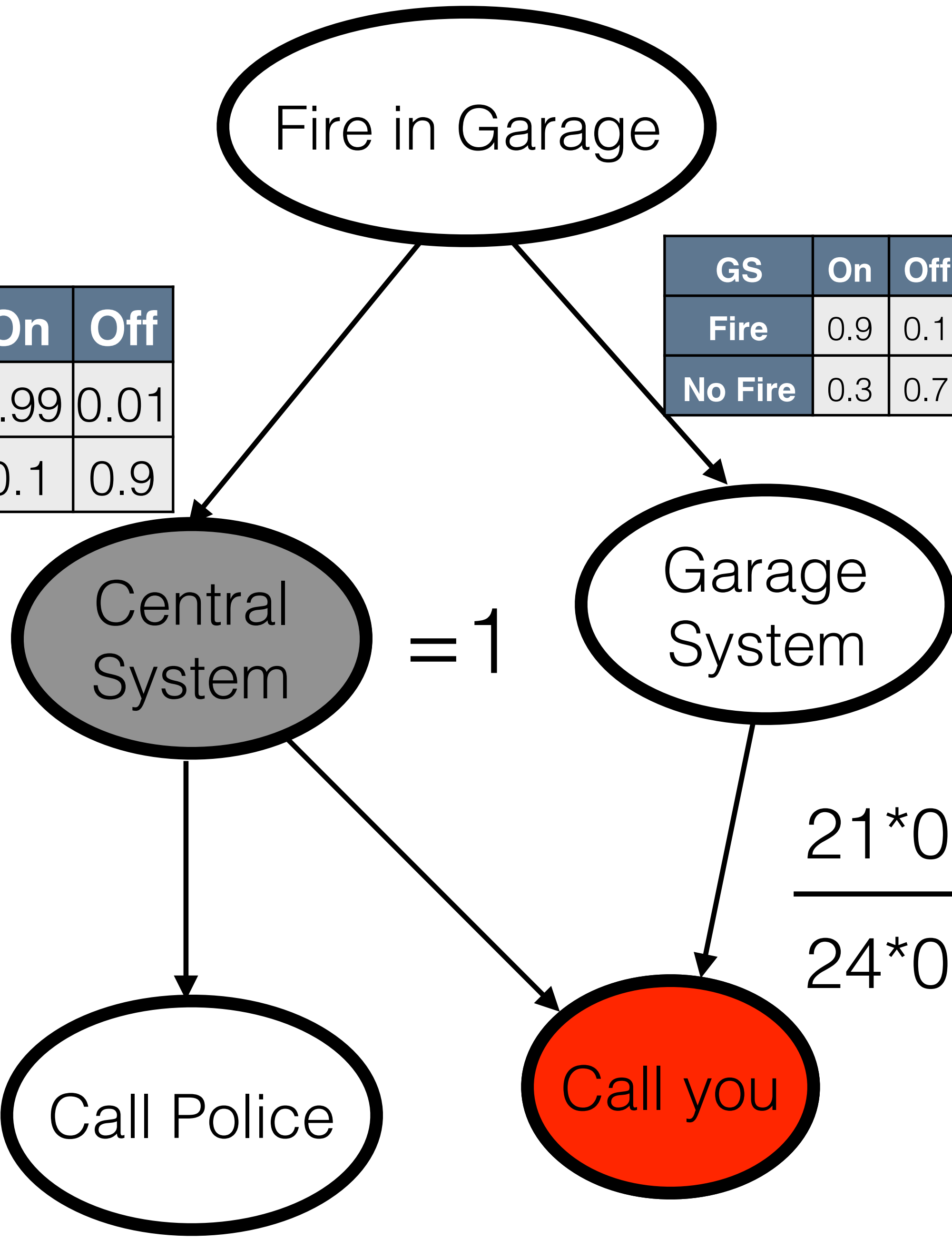
P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no
CS	GS	0.95	0.05
CS	GS	0.7	0.3
CS	GS	0.9	0.1
CS	GS	0.01	0.99

	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	
2	0	1	1	1	1	
3	1	1	1	1	1	
4	0	1	0	1	1	
5	0	1	0	1	1	
6	0	1	0	1	1	
7	0	1	1	1	1	
8	0	1	0	1	0	
9	0	1	0	1	1	
[illegible]	0	1	1	1	1	
[illegible]	0	1	0	1	1	
[illegible]	0	1	1	1	0	
13	0	1	0	1	1	
14	0	1	1	1	1	
15	0	1	0	1	1	
16	0	1	0	1	1	
17	0	1	0	1	1	
18	0	1	0	1	1	
19	0	1	1	1	1	
20	0	1	0	1	1	
21	0	1	0	1	1	
22	0	1	0	1	1	
23	0	1	1	1	1	
24	0	1	0	1	1	
25	0	1	1	1	1	

Importance Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

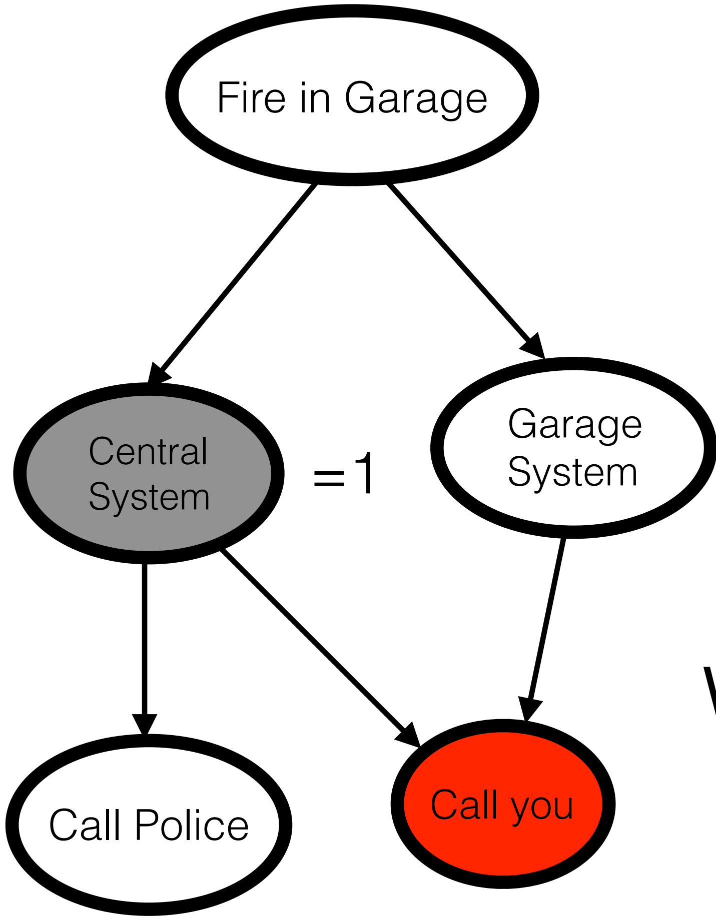
$$\frac{21*0.1 + 1\times 0.99}{24*0.1 + 1\times 0.99}$$

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no
CS	GS	0.95	0.05
CS	GS	0.7	0.3
CS	GS	0.9	0.1
CS	GS	0.01	0.99

	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	0.1
2	0	1	1	1	1	0.1
3	1	1	1	1	1	0.99
4	0	1	0	1	1	0.1
5	0	1	0	1	1	0.1
6	0	1	0	1	1	0.1
7	0	1	1	1	1	0.1
8	0	1	0	1	0	0.1
9	0	1	0	1	1	0.1
10	0	1	1	1	1	0.1
11	0	1	0	1	1	0.1
12	0	1	1	1	0	0.1
13	0	1	0	1	1	0.1
14	0	1	1	1	1	0.1
15	0	1	0	1	1	0.1
16	0	1	0	1	1	0.1
17	0	1	0	1	1	0.1
18	0	1	0	1	1	0.1
19	0	1	1	1	1	0.1
20	0	1	0	1	1	0.1
21	0	1	0	1	1	0.1
22	0	1	0	1	1	0.1
23	0	1	1	1	1	0.1
24	0	1	0	1	1	0.1
25	0	1	1	1	1	0.1

Importance Sampling

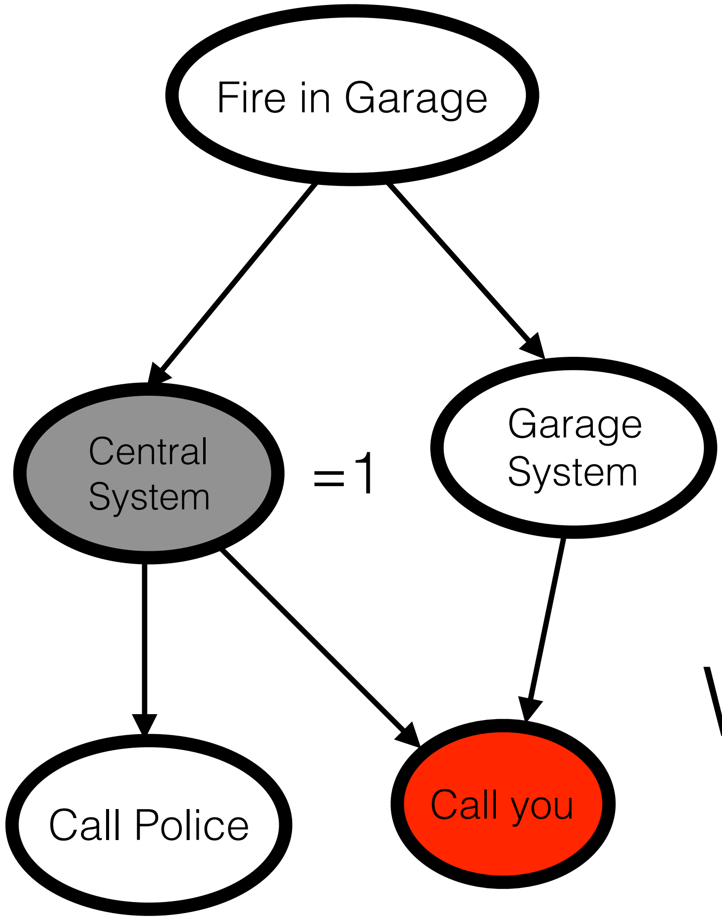


	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

What we want: Draw from P(F,GS,P,Y| CS=1)

$$P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)$$

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

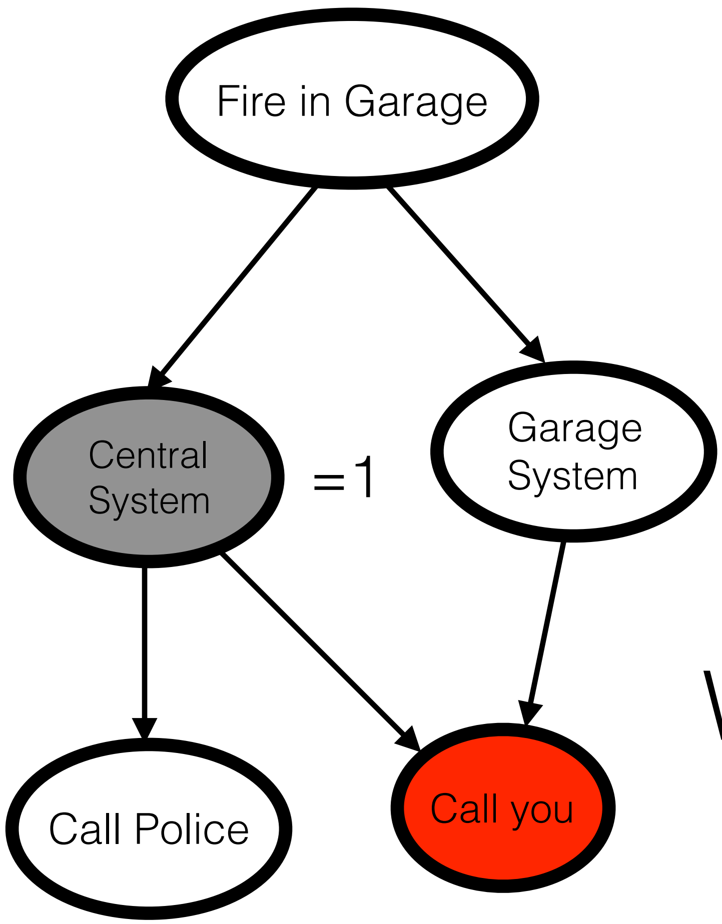
What we want: Draw from P(F,GS,P,Y| CS=1)

$$P(\mathrm{F}=0, \mathrm{GS}=0, \mathrm{P}=1, \mathrm{Y}=0 | \mathrm{CS}=1)$$

Instead we draw from ?

$$P(\mathrm{F}=0)$$

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

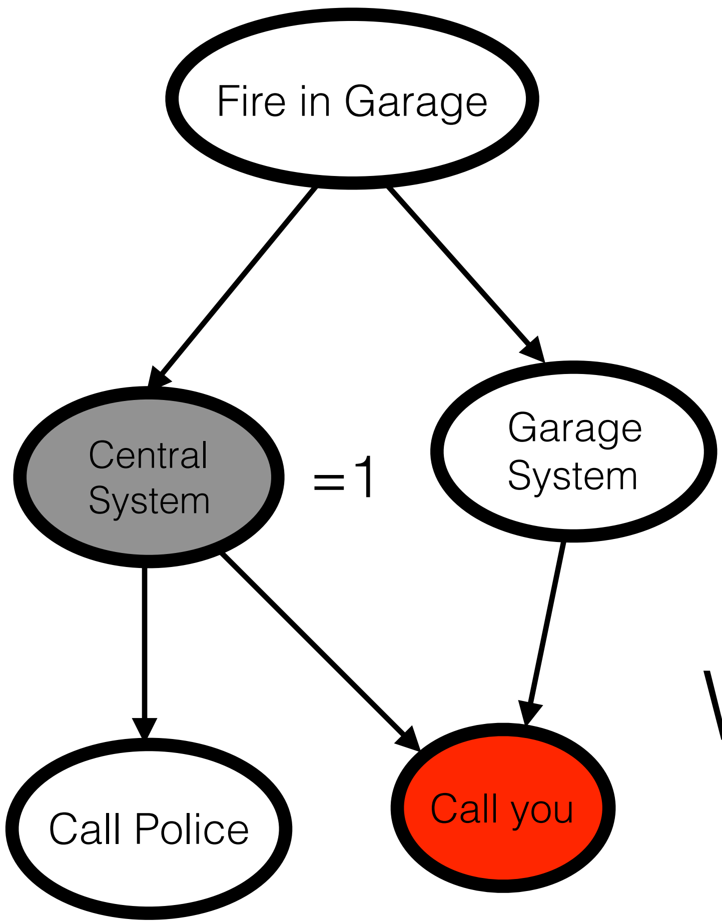
What we want: Draw from P(F,GS,P,Y| CS=1)

$$P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)$$

Instead we draw from ?

$$P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)$$

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

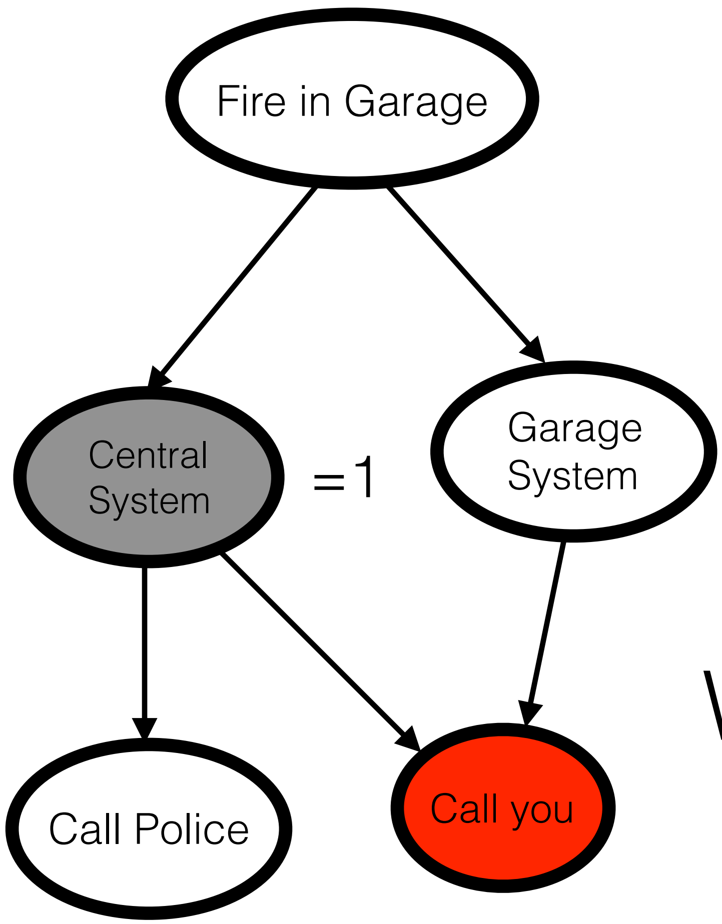
What we want: Draw from P(F,GS,P,Y| CS=1)

$$P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)$$

Instead we draw from ?

$$P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)\times P(\mathrm{P}=1|\mathrm{CS}=1)$$

IMPORTANCE SAMPLING



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

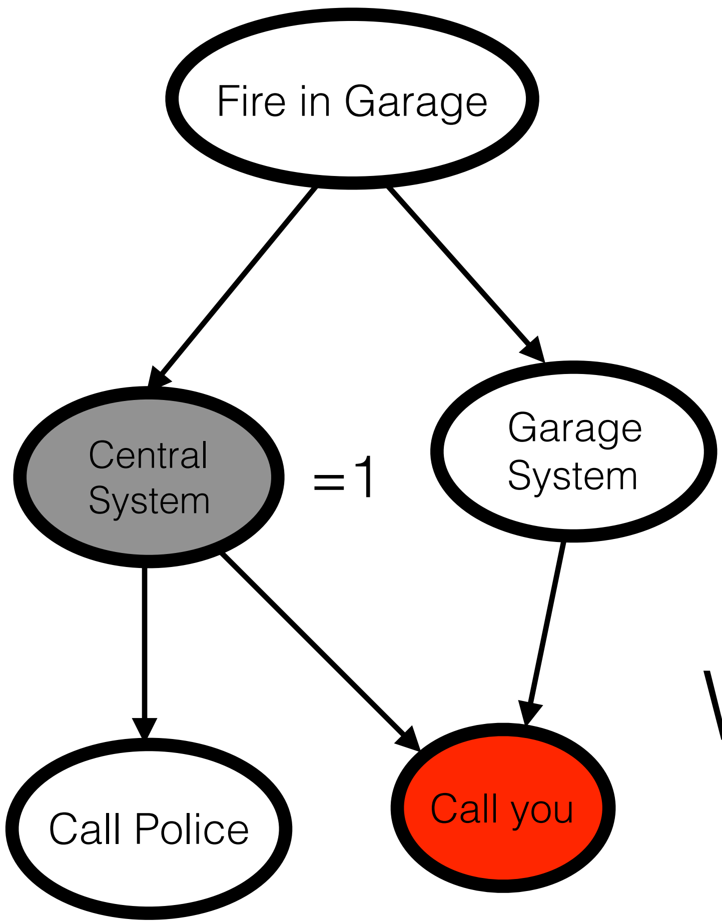
What we want: Draw from P(F,GS,P,Y| CS=1)

$$P(\mathrm{F}=0, \mathrm{GS}=0, \mathrm{P}=1, \mathrm{Y}=0 | \mathrm{CS}=1)$$

Instead we draw from ?

$$P(\mathrm{F}=0) \times P(\mathrm{GS}=0 | \mathrm{F}=0) \times P(\mathrm{P}=1 | \mathrm{CS}=1) \times P(\mathrm{Y}=0 | \mathrm{CS}=1, \mathrm{GS}=0)$$

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

What we want: Draw from P(F,GS,P,Y| CS=1)

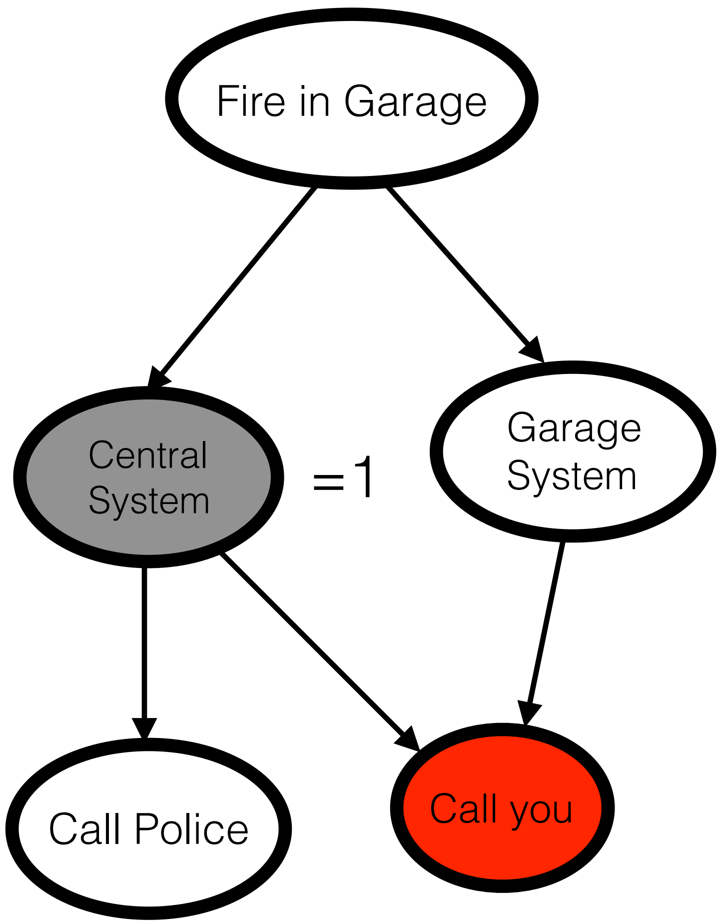
$$P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)$$

Instead we draw from ?

$$P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)\times P(\mathrm{P}=1|\mathrm{CS}=1)\times P(\mathrm{Y}=0|\mathrm{CS}=1,\mathrm{GS}=0)$$

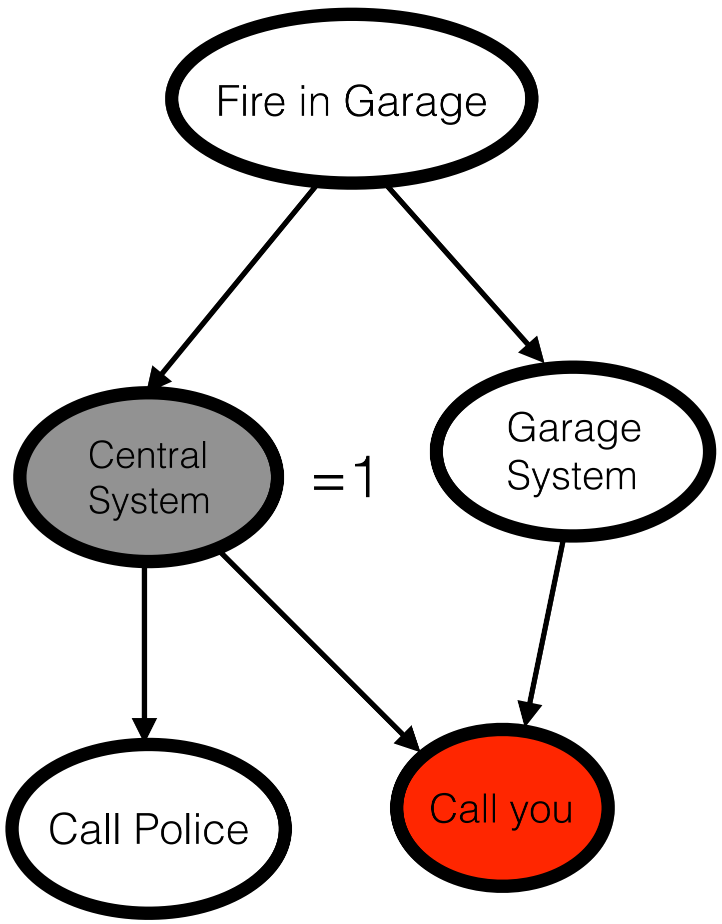
Weigh each sample by ratio of Prob we want / Prob of draw

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

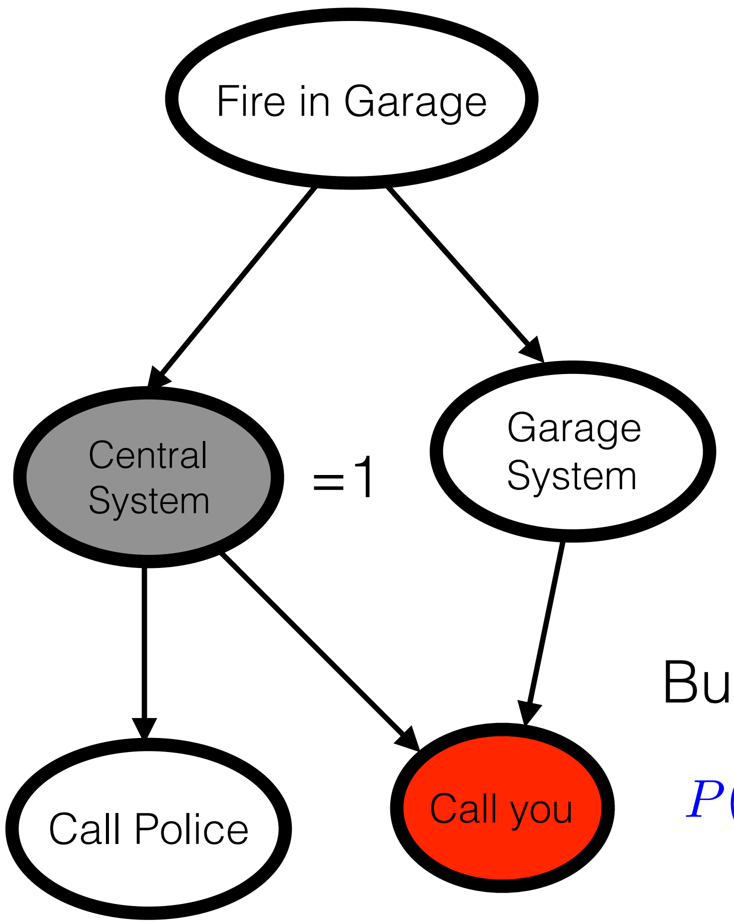
Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

$$\text{weight} \propto \frac{P(\mathrm{F}=0, \mathrm{GS}=0, \mathrm{P}=1, \mathrm{Y}=0 | \mathrm{CS}=1)}{P(\mathrm{F}=0) \times P(\mathrm{GS}=0 | \mathrm{F}=0) \times P(\mathrm{P}=1 | \mathrm{CS}=1) \times P(\mathrm{Y}=1 | \mathrm{GS}=0, \mathrm{CS}=1)}$$

IMPORTANCE SAMPLING



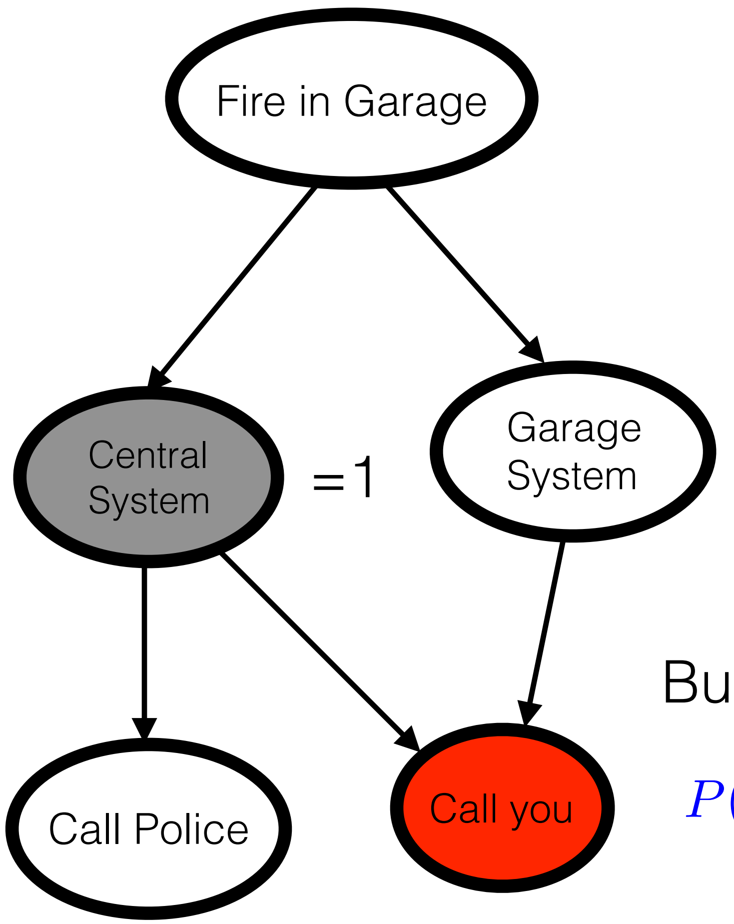
	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

But:

$$P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)=\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0,\mathrm{CS}=1)}{P(\mathrm{CS}=1)}$$

$$\mathrm{weight}\propto\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)}{P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)\times P(\mathrm{P}=1|\mathrm{CS}=1)\times P(\mathrm{Y}=1|\mathrm{GS}=0,\mathrm{CS}=1)}$$

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

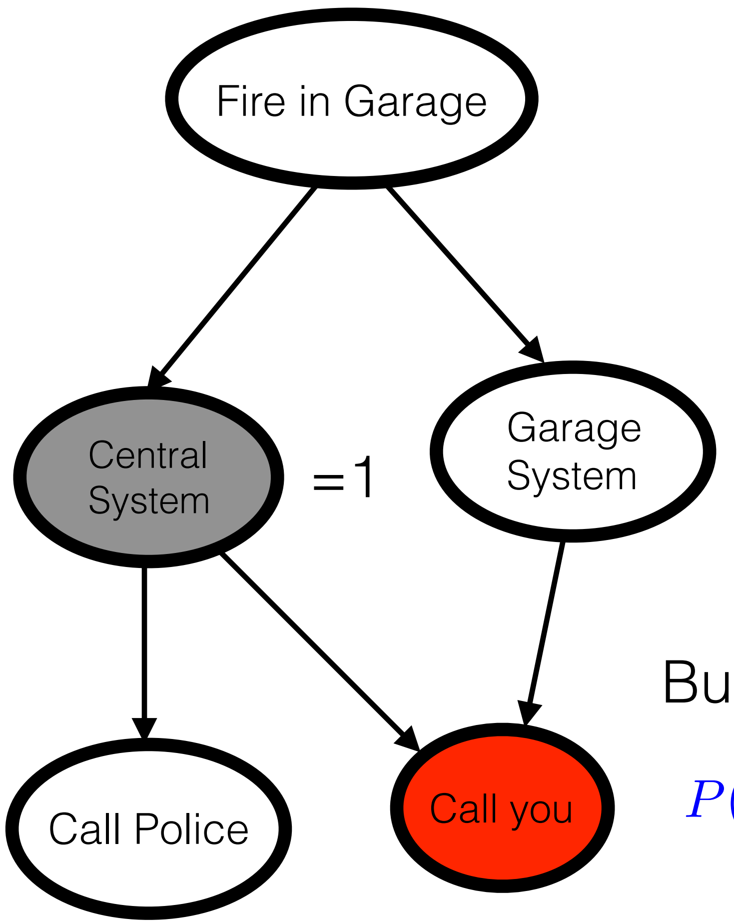
But:

$$P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)=\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0,\mathrm{CS}=1)}{P(\mathrm{CS}=1)}$$

$$\text{weight}\propto\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)}{P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)\times P(\mathrm{P}=1|\mathrm{CS}=1)\times P(\mathrm{Y}=1|\mathrm{GS}=0,\mathrm{CS}=1)}$$

$$\propto\frac{1}{P(\mathrm{CS}=1)}\cdot\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0,\mathrm{CS}=1)}{P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)\times P(\mathrm{P}=1|\mathrm{CS}=1),P(\mathrm{Y}=0|\mathrm{CS}=1,\mathrm{GS}=0)}$$

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

But:

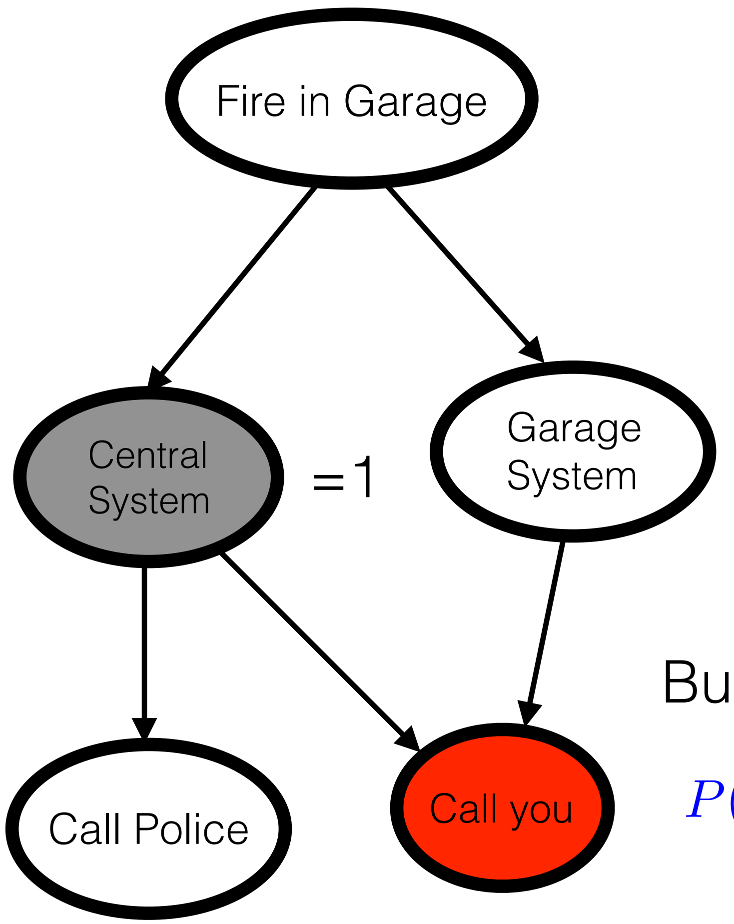
$$P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)=\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0,\mathrm{CS}=1)}{P(\mathrm{CS}=1)}$$

$$\text{weight}\propto\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0|\mathrm{CS}=1)}{P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)\times P(\mathrm{P}=1|\mathrm{CS}=1)\times P(\mathrm{Y}=1|\mathrm{GS}=0,\mathrm{CS}=1)}$$

$$\propto\frac{1}{P(\mathrm{CS}=1)}\cdot\frac{P(\mathrm{F}=0,\mathrm{GS}=0,\mathrm{P}=1,\mathrm{Y}=0,\mathrm{CS}=1)}{P(\mathrm{F}=0)\times P(\mathrm{GS}=0|\mathrm{F}=0)\times P(\mathrm{P}=1|\mathrm{CS}=1),P(\mathrm{Y}=0|\mathrm{CS}=1,\mathrm{GS}=0)}$$

$$\propto\frac{P(\mathrm{CS}=1|\mathrm{F}=0)}{P(\mathrm{CS}=1)}$$

Importance Sampling



	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	

But:

$$P(\mathrm{F}=0, \mathrm{GS}=0, \mathrm{P}=1, \mathrm{Y}=0 | \mathrm{CS}=1) = \frac{P(\mathrm{F}=0, \mathrm{GS}=0, \mathrm{P}=1, \mathrm{Y}=0, \mathrm{CS}=1)}{P(\mathrm{CS}=1)}$$

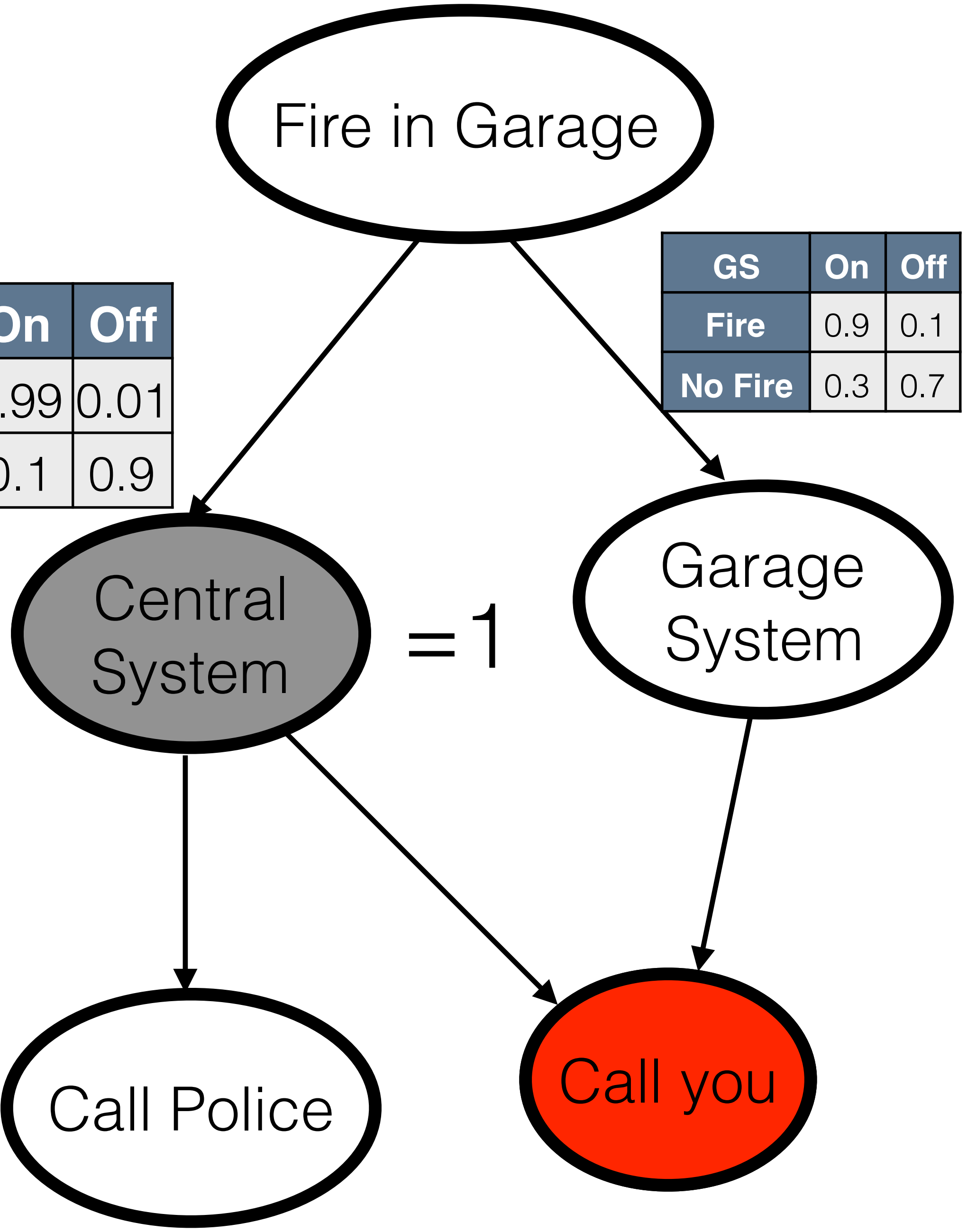
$$\text{weight} \propto \frac{P(\mathrm{F}=0, \mathrm{GS}=0, \mathrm{P}=1, \mathrm{Y}=0 | \mathrm{CS}=1)}{P(\mathrm{F}=0) \times P(\mathrm{GS}=0|\mathrm{F}=0) \times P(\mathrm{P}=1|\mathrm{CS}=1) \times P(\mathrm{Y}=1|\mathrm{GS}=0, \mathrm{CS}=1)}$$

$$\propto \frac{1}{P(\mathrm{CS}=1)} \cdot \frac{P(\mathrm{F}=0, \mathrm{GS}=0, \mathrm{P}=1, \mathrm{Y}=0, \mathrm{CS}=1)}{P(\mathrm{F}=0) \times P(\mathrm{GS}=0|\mathrm{F}=0) \times P(\mathrm{P}=1|\mathrm{CS}=1), P(\mathrm{Y}=0|\mathrm{CS}=1, \mathrm{GS}=0)}$$

$$\propto \frac{P(\mathrm{CS}=1|\mathrm{F}=0)}{P(\mathrm{CS}=1)} \quad \propto P(\mathrm{CS}=1|\mathrm{F}=0)$$

Importance Sampling

Fire	No Fire
0.05	0.95



CS	On	Off
Fire	0.99	0.01
No Fire	0.1	0.9

GS	On	Off
Fire	0.9	0.1
No Fire	0.3	0.7

P	Call	no call
CS on	1	0
CS off	0.01	0.99

Y		Call	no
CS	GS	0.95	0.05
CS	GS	0.7	0.3
CS	GS	0.9	0.1
CS	GS	0.01	0.99

	F	CS	GS	P	Y	Weight
1	0	1	0	1	0	0.1
2	0	1	1	1	1	0.1
3	1	1	1	1	1	0.99
4	0	1	0	1	1	0.1
5	0	1	0	1	1	0.1
6	0	1	0	1	1	0.1
7	0	1	1	1	1	0.1
8	0	1	0	1	0	0.1
9	0	1	0	1	1	0.1
10	0	1	1	1	1	0.1
11	0	1	0	1	1	0.1
12	0	1	1	1	0	0.1
13	0	1	0	1	1	0.1
14	0	1	1	1	1	0.1
15	0	1	0	1	1	0.1
16	0	1	0	1	1	0.1
17	0	1	0	1	1	0.1
18	0	1	0	1	1	0.1
19	0	1	1	1	1	0.1
20	0	1	0	1	1	0.1
21	0	1	0	1	1	0.1
22	0	1	0	1	1	0.1
23	0	1	1	1	1	0.1
24	0	1	0	1	1	0.1
25	0	1	1	1	1	0.1

IMPORTANCE SAMPLING

Likelihood weighting:

Topologically sort variables (parents first children later)

For $t = 1$ to n (number of samples)

Set $w_t = 1$

For $i = 1$ to N (number of variables)

If X_i is observed,

Set $w_t \leftarrow w_t \cdot P(X_i = x_i | \text{Parents}(X_i) = \text{ already sampled})$

Set $x_i^t = x_i$ (the observed value)

Else, sample $x_i^t \sim P(X_i | \text{Parents}(X_i) = \text{ already sampled})$

End For

End For

Output,

$$P(\text{Variable} = \text{value} | \text{Observation}) = \frac{\sum_{t=1}^{n} w_t \mathbf{1}\{\text{Variable} = \text{value}\}}{\sum_{t=1}^{n} w_t}$$

Importance Sampling

- We really want to draw from distribution P.
- But we can only draw from distribution Q easily
- Trick:
  - Draw $x_1, \ldots, x_n \sim Q$
  - Re-weight each sample x_t by $P(X = x_t)/Q(X = x_t)$

Importance Sampling

- Why does it work?

$$\mathbb{E}_{X\sim P}[f(X)] = \sum_x P(X = x)f(x)$$

Importance Sampling

- Why does it work?

$$\mathbb{E}_{X \sim P}[f(X)] = \sum_x P(X = x) f(x)$$
$$= \sum_x Q(X = x) \left(\frac{P(X = x)}{Q(X = x)} f(x) \right)$$

Importance Sampling

- Why does it work?

$$\begin{aligned}\mathbb{E}_{X\sim P}[f(X)] &= \sum_x P(X=x)f(x)\\ &= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)}f(x)\right)\\ &= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)}f(X)\right]\end{aligned}$$

Importance Sampling

- Why does it work?

$$\begin{aligned}
\mathbb{E}_{X\sim P}[f(X)] &= \sum_x P(X=x)f(x) \\
&= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)}f(x)\right) \\
&= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)}f(X)\right] \\
&\approx \frac{1}{n}\sum_{t=1}^{n}\frac{P(X=x_t)}{Q(X=x_t)}f(x_t)
\end{aligned}$$

Importance Sampling

- Why does it work?

$$
\begin{aligned}
\mathbb{E}_{X\sim P}[f(X)] &= \sum_x P(X=x) f(x) \\
&= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)} f(x)\right) \\
&= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)} f(X)\right] \\
&\approx \frac{1}{n}\sum_{t=1}^{n} \frac{P(X=x_t)}{Q(X=x_t)} f(x_t)
\end{aligned}
$$

- Example: $f(X) = \mathbf{1}\{X \in \text{Set}\}$, then $\mathbb{E}_{X\sim P}[f(X)] = P(X \in \text{Set})$

Importance Sampling

- Why does it work?

$$\begin{aligned}
\mathbb{E}_{X\sim P}[f(X)] &= \sum_x P(X=x)f(x) \\
&= \sum_x Q(X=x)\left(\frac{P(X=x)}{Q(X=x)}f(x)\right) \\
&= \mathbb{E}_{X\sim Q}\left[\frac{P(X)}{Q(X)}f(X)\right] \\
&\approx \frac{1}{n}\sum_{t=1}^{n}\frac{P(X=x_t)}{Q(X=x_t)}f(x_t)
\end{aligned}$$

- Example: $f(X) = \mathbf{1}\{X \in \text{Set}\}$, then $\mathbb{E}_{X\sim P}[f(X)] = P(X \in \text{Set})$
- Hence, using importance weighted sampling,

$$P(X \in \text{Set}) \approx \frac{1}{n}\sum_{t=1}^{n}\mathbf{1}\{x_t \in \text{Set}\}\frac{P(X=x_t)}{Q(X=x_t)}$$