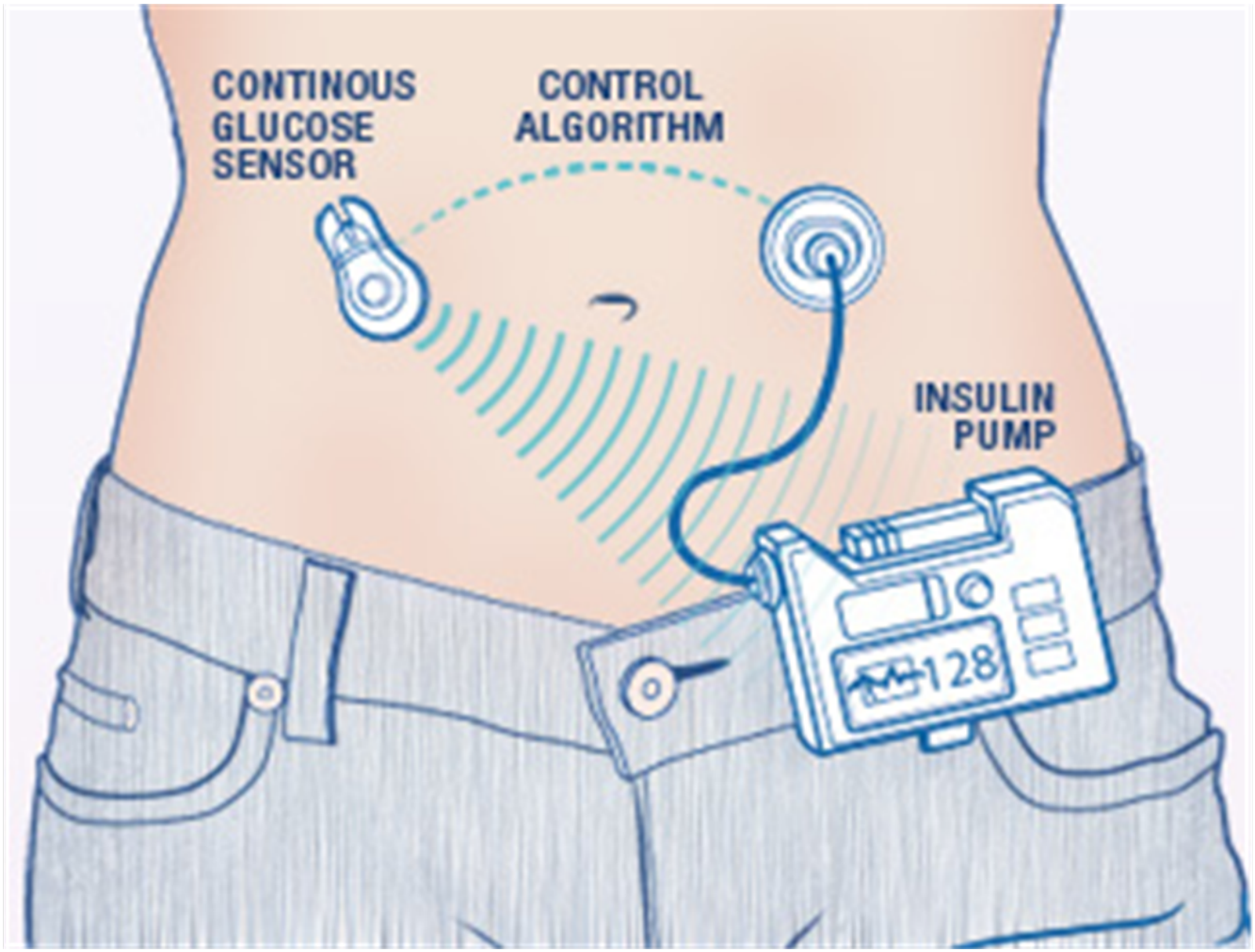
Causal Inference under Uncertainty via Adjustments and SOPDs

Dylan Hutchison and Samantha Kleinberg
Stevens Institute of Technology


94
M


Medtronic
MiniMed
2:33PM
2:33P
24 Hour
108
PARADIGM
B
ESC
ACT


CONTINOUS
GLUCOSE
SENSOR
CONTROL
ALGORITHM
INSULIN
PUMP
128

Time	Blood Glucose	Heart Rate	Skin Temp.	Energy Use	Weight
7:20					
7:25		95.7	26.80	**37.23**	
7:30		90.5	28.63	12.85	
7:35		82.4	29.69	6.78	
7:40		79.2	30.32	6.41	
7:45		77.0	30.55	8.17	
7:50		73.3	31.19	6.87	**155**
7:55		72.7	31.29	6.08	
8:00	**130.0**	71.4	31.96	9.74	
8:05	203.5	89.8	31.59	24.58	
8:10	208.2	87.9	31.67	10.63	
8:15	209.0	84.8	31.16	15.63	
8:20	207.8	81.0	31.11	11.74	
8:25	205.9	83.7	31.26	16.28	
8:30	204.8	98.3	31.05	23.59	
8:35	214.9	82.2	30.47	11.66	
8:40	225.7	82.9	30.95	16.93	
8:45	231.4	**0.1**			
8:50	232.8	89.6	29.67	21.96	
8:55	239.4	81.2	30.81	15.86	

Time	Blood Glucose	Heart Rate	Skin Temp.	Energy Use	Weight
7:20					
7:25		95.7	26.80	**37.23**	
7:30		90.5	28.63	12.85	
7:35		82.4	29.69	6.78	
7:40		79.2	30.32	6.41	
7:45		77.0	30.55	8.17	
7:50		73.3	31.19	6.87	**155**
7:55		72.7	31.29	6.08	
8:00	**130.0**	71.4	31.96	9.74	
8:05	203.5	89.8	31.59	24.58	
8:10	208.2	87.9	31.67	10.63	
8:15	209.0	84.8	31.16	15.63	
8:20	207.8	81.0	31.11	11.74	
8:25	205.9	83.7	31.26	16.28	
8:30	204.8	98.3	31.05	23.59	
8:35	214.9	82.2	30.47	11.66	
8:40	225.7	82.9	30.95	16.93	
8:45	231.4	**0.1**			
8:50	232.8	89.6	29.67	21.96	
8:55	239.4	81.2	30.81	15.86	

Incorrect Data

Time	Blood Glucose	Heart Rate	Skin Temp.	Energy Use	Weight
7:20					
7:25		95.7	26.80	**37.23**	
7:30		90.5	28.63	12.85	
7:35		82.4	29.69	6.78	
7:40		79.2	30.32	6.41	
7:45		77.0	30.55	8.17	
7:50		73.3	31.19	6.87	**155**
7:55		72.7	31.29	6.08	
8:00	**130.0**	71.4	31.96	9.74	
8:05	203.5	89.8	31.59	24.58	
8:10	208.2	87.9	31.67	10.63	
8:15	209.0	84.8	31.16	15.63	
8:20	207.8	81.0	31.11	11.74	
8:25	205.9	83.7	31.26	16.28	
8:30	204.8	98.3	31.05	23.59	
8:35	214.9	82.2	30.47	11.66	
8:40	225.7	82.9	30.95	16.93	
8:45	231.4	**0.1**			
8:50	232.8	89.6	29.67	21.96	
8:55	239.4	81.2	30.81	15.86	

Incorrect Data

Incomplete Data

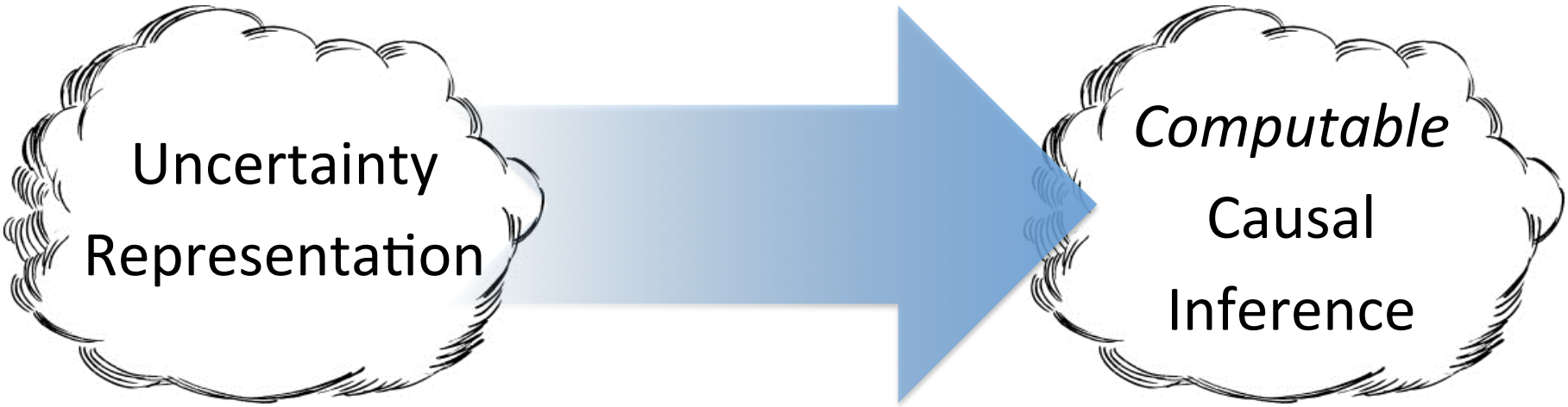
Time	Blood Glucose	Heart Rate	Skin Temp.	Energy Use	Weight
7:20					
7:25		95.7	26.80	**37.23**	
7:30		90.5	28.63	12.85	
7:35		82.4	29.69	6.78	
7:40		79.2	30.32	6.41	
7:45		77.0	30.55	8.17	
7:50		73.3	31.19	6.87	**155**
7:55		72.7	31.29	6.08	
8:00	**130.0**	71.4	31.96	9.74	
8:05	203.5	89.8	31.59	24.58	
8:10	208.2	87.9	31.67	10.63	
8:15	209.0	84.8	31.16	15.63	
8:20	207.8	81.0	31.11	11.74	
8:25	205.9	83.7	31.26	16.28	
8:30	204.8	98.3	31.05	23.59	
8:35	214.9	82.2	30.47	11.66	
8:40	225.7	82.9	30.95	16.93	
8:45	231.4	**0.1**			
8:50	232.8	89.6	29.67	21.96	
8:55	239.4	81.2	30.81	15.86	

Incorrect Data

Incomplete Data

Outdated Data

What's missing?



Intervals

Belief Functions

Base Weights

Granger VAR

Temporal Logic

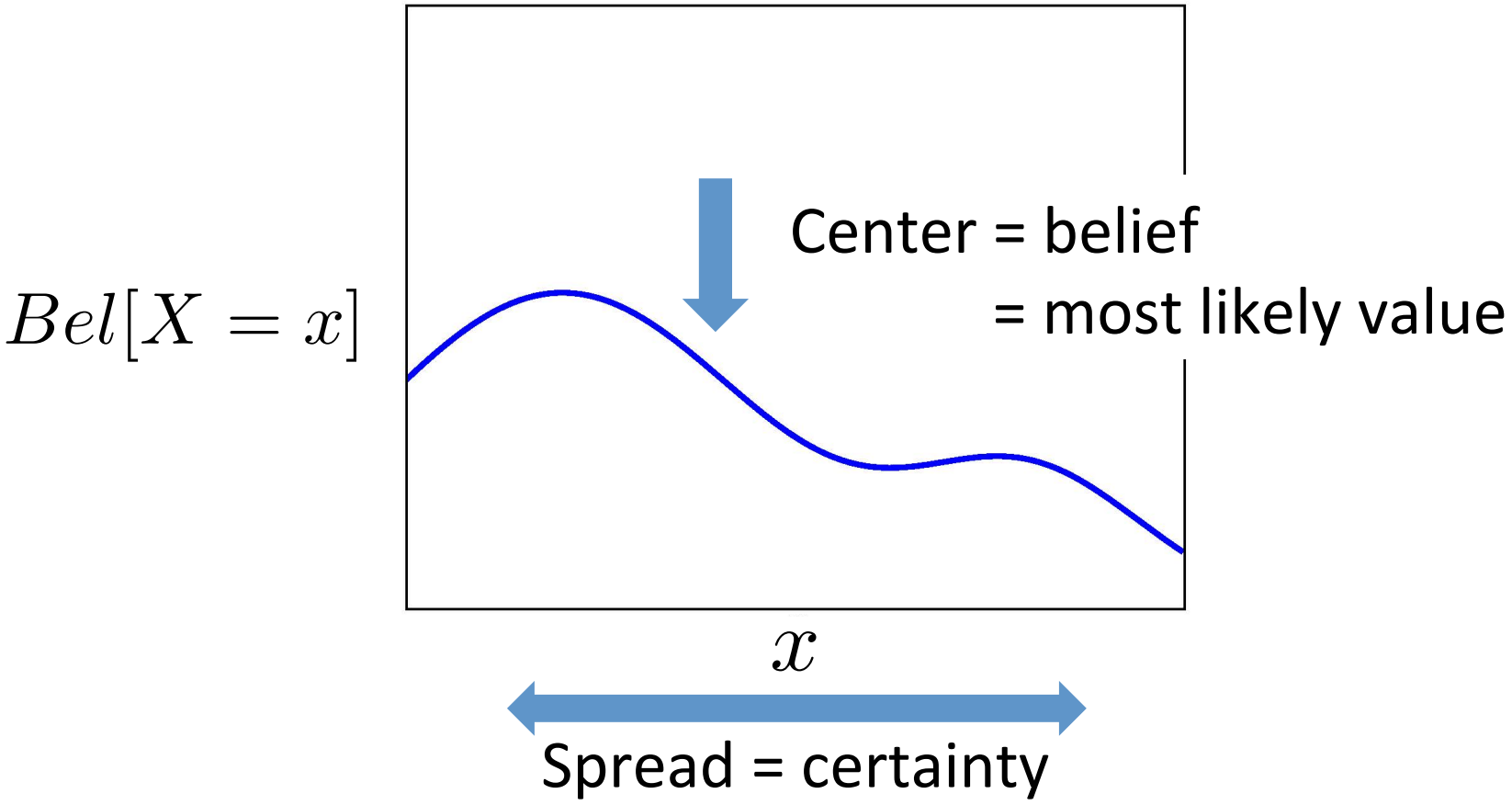
DBNs

Approach

Assign uncertain values less weight

- **Inaccurate Data** – assign less weight
- **Missing Data** – impute with less weight
- **Outdated Data** – weigh with decay factor

Second Order Probability Distribution



(center, spread)

Time	Blood Glucose	Skin Temp.	Energy Use	Weight
7:20				
7:25		26.8	**37.2**	
7:30		28.6	12.9	
7:35		29.7	6.8	
7:40		30.3	6.4	
7:45		30.5	8.2	
7:50		31.2	6.9	155
7:55		31.3	6.1	
8:00	**186.0**	32.0	9.7	
8:05	253.5	31.6	24.6	
8:10	258.2	31.7	10.6	

Time	Blood Glucose	Skin Temp.	Energy Use	Weight
7:20				
7:25		(26.8, 0.6)	**(37.2, 0.6)**	
7:30		(28.6, 0.8)	(12.9, 0.8)	
7:35		(29.7,1)	(6.8,1)	
7:40		(30.3,1)	(6.4,1)	
7:45		(30.6,1)	(8.2,1)	
7:50		(31.2,1)	(6.9,1)	(155,1)
7:55		(31.3,1)	(6.1,1)	
8:00	**(186.0, 0.6)**	(32.0,1)	(9.7,1)	
8:05	(253.5, 0.8)	(31.6,1)	(24.6,1)	
8:10	(258.2,1)	(31.7,1)	(10.6,1)	

Discretization

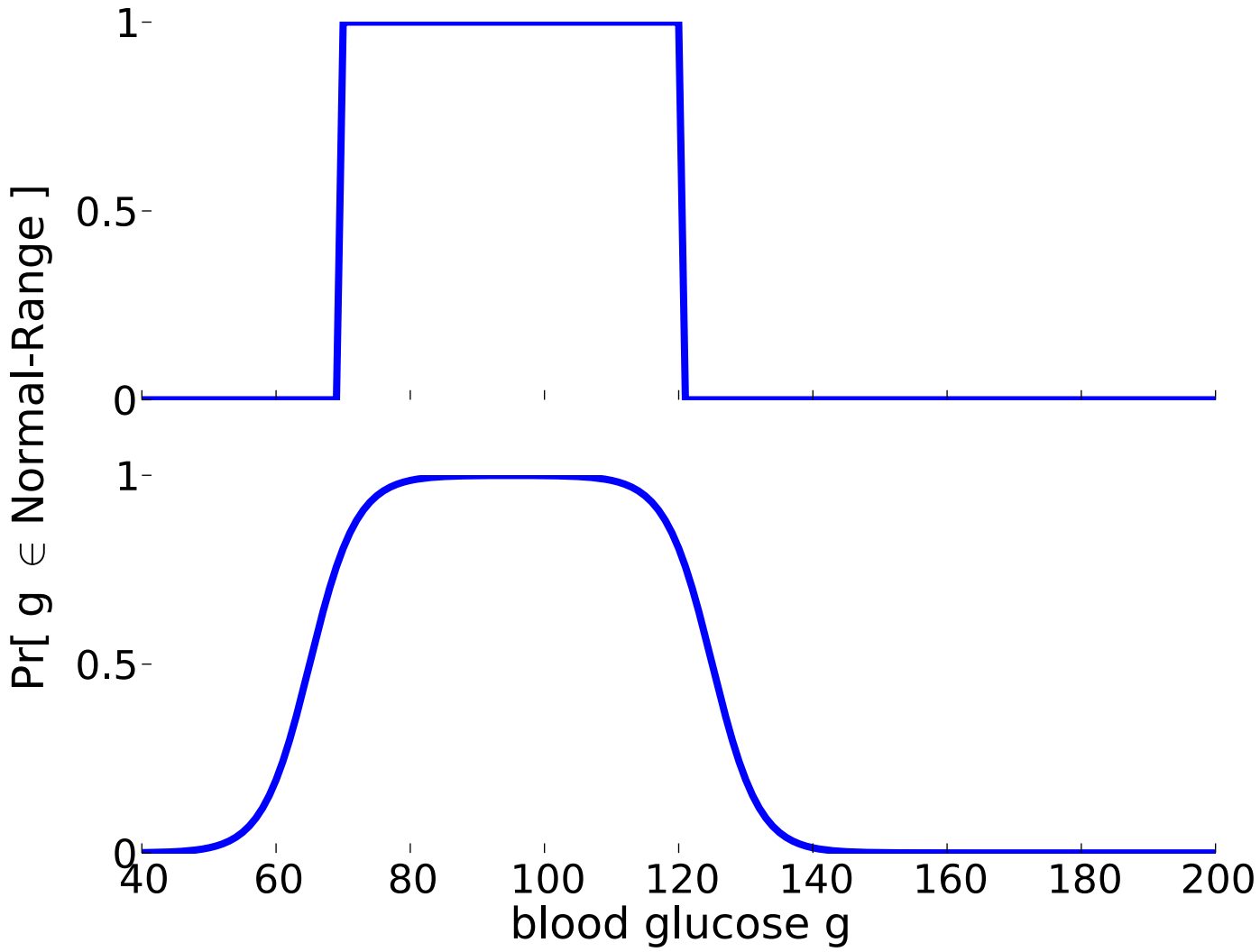
- How do we partition continuous data into discrete events?

Time	Glucose
8:00	130

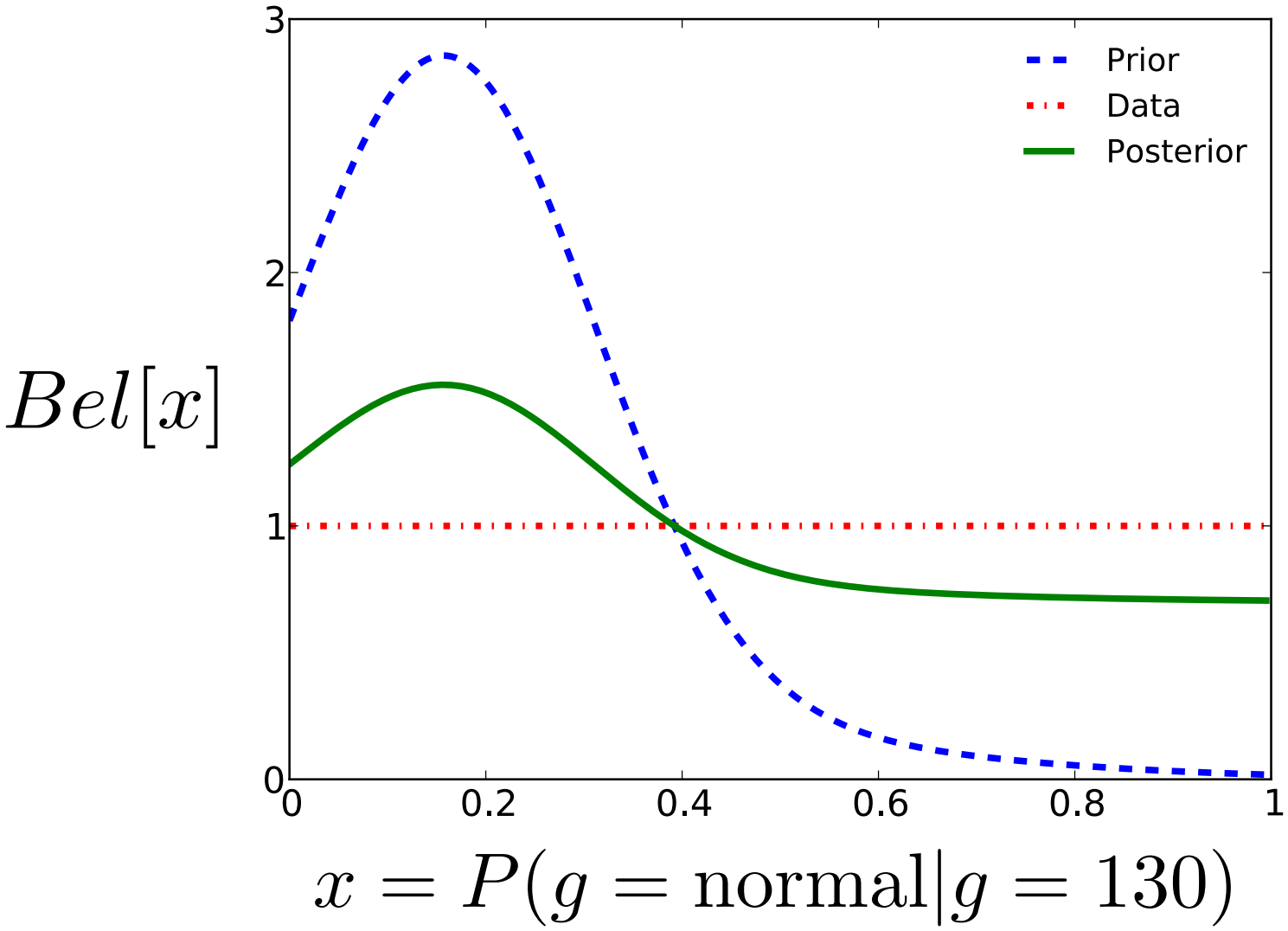
→ {too low, normal, too high} ?

- Strict partitioning leads to inconsistencies

Uncertainty in discretization



SOPD Adiustment



Causal inference, without uncertainty

- Complex, temporal relationships

$$v \overset{\geq 15, \leq 40}{\underset{\geq 0.4}{\rightsquigarrow}} g$$

- Assess average difference cause makes to probability of effect

$$\varepsilon_{avg}(c, e) = \frac{\sum_{x \in X \ c} P(e|c \wedge x) - P(e|\neg c \wedge x)}{|X \backslash c|}$$

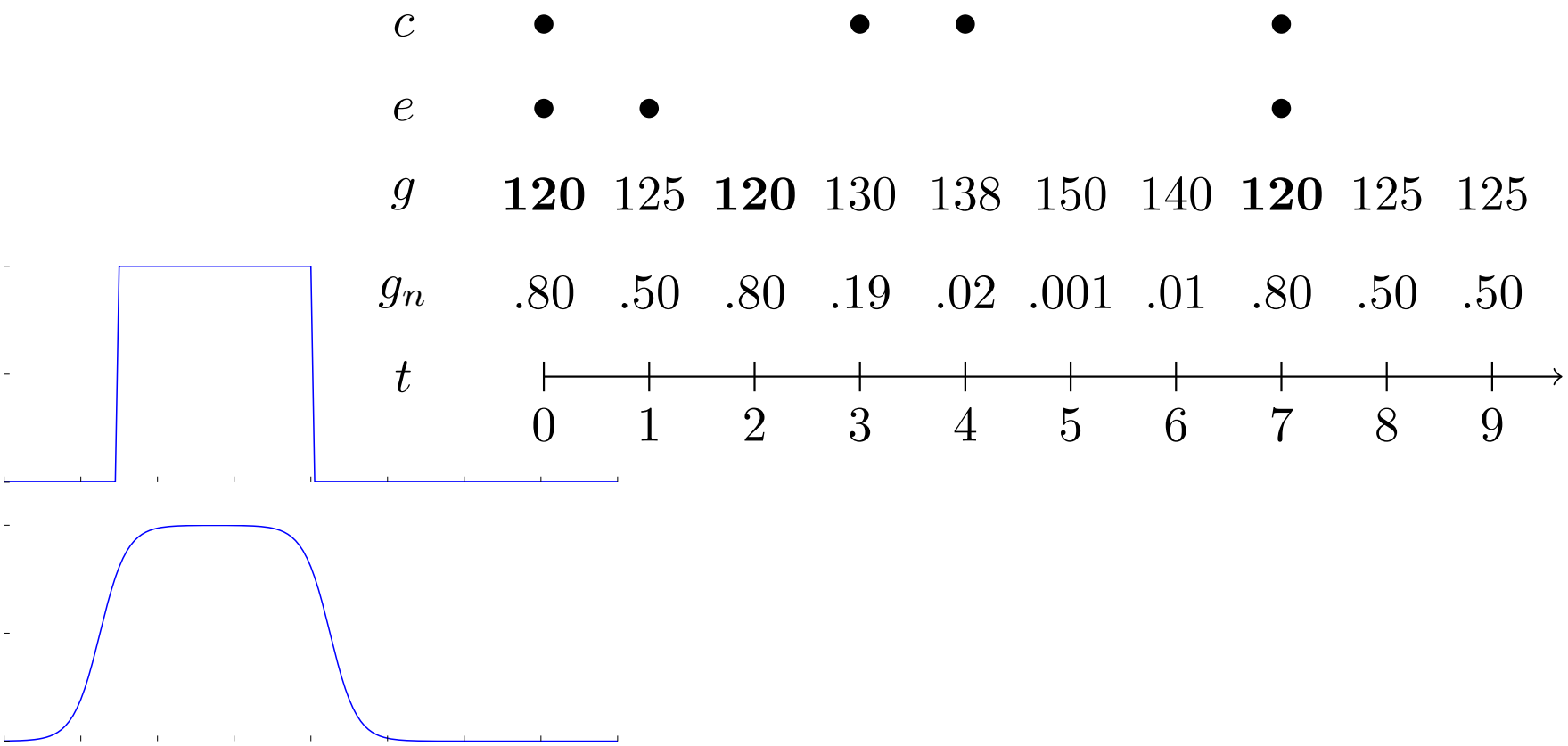
Kleinberg, S. (2013) Causality, Probability, and Time.

Adding uncertainty

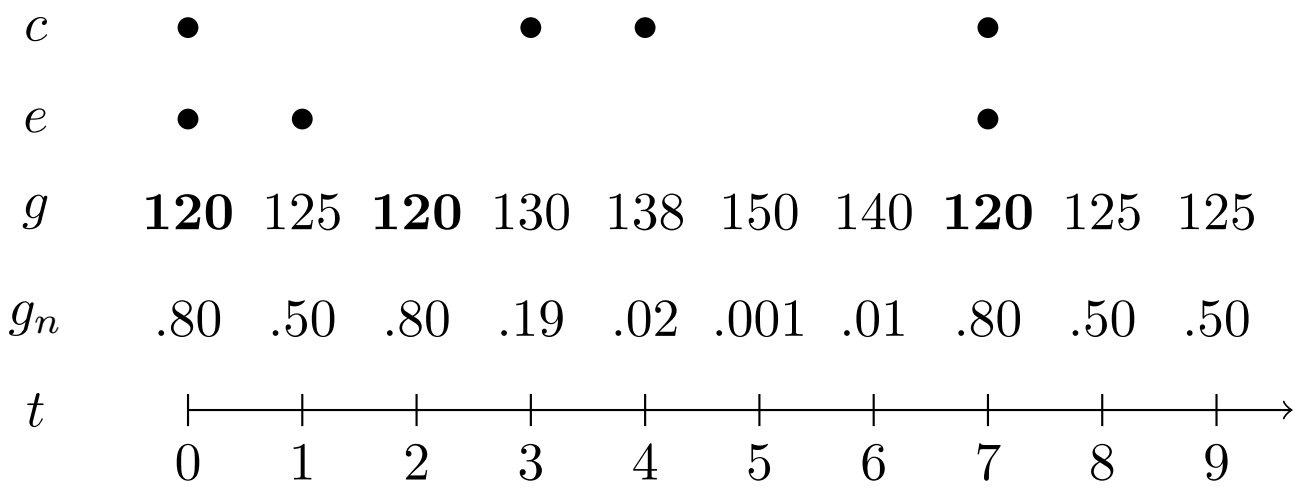
$$P(e|c,x) = \frac{\sum_t ecx}{\sum_t cx}$$

$$P(e|c,x) = \frac{\sum_t e^c e^s c^c c^s x^c x^s}{\sum_t e^s c^c c^s x^c x^s}$$

Carb-heavy meal (c), vigorous exercise (e), blood glucose (g)



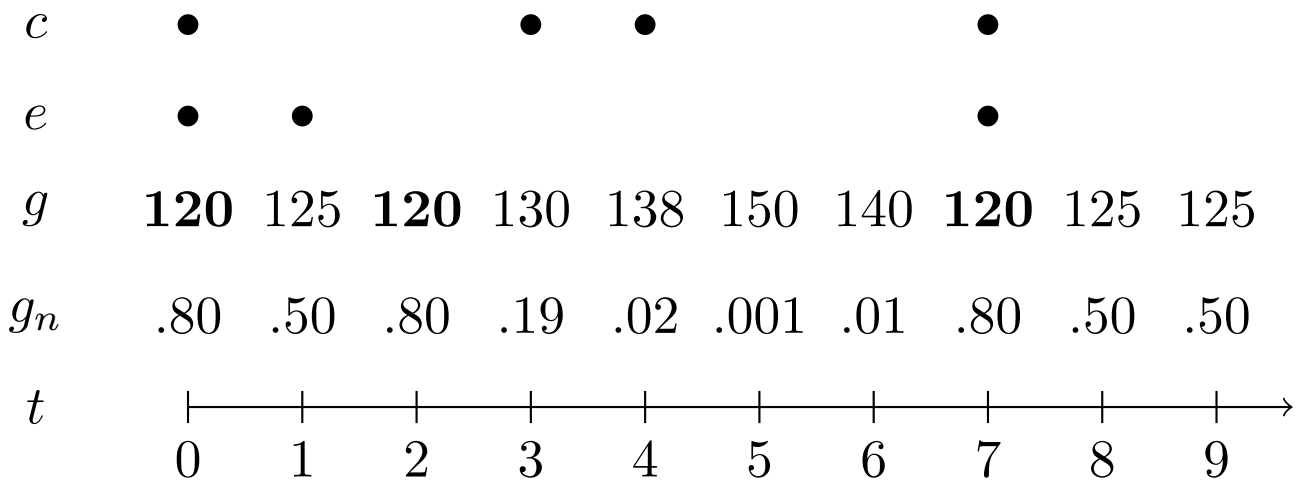
What's the effect of exercise on glucose?



$$e \rightsquigarrow^{\geq 1, \leq 1} g_n$$

Calculate: $P(g_n|c, e) - P(g_n|c, \neg e)$

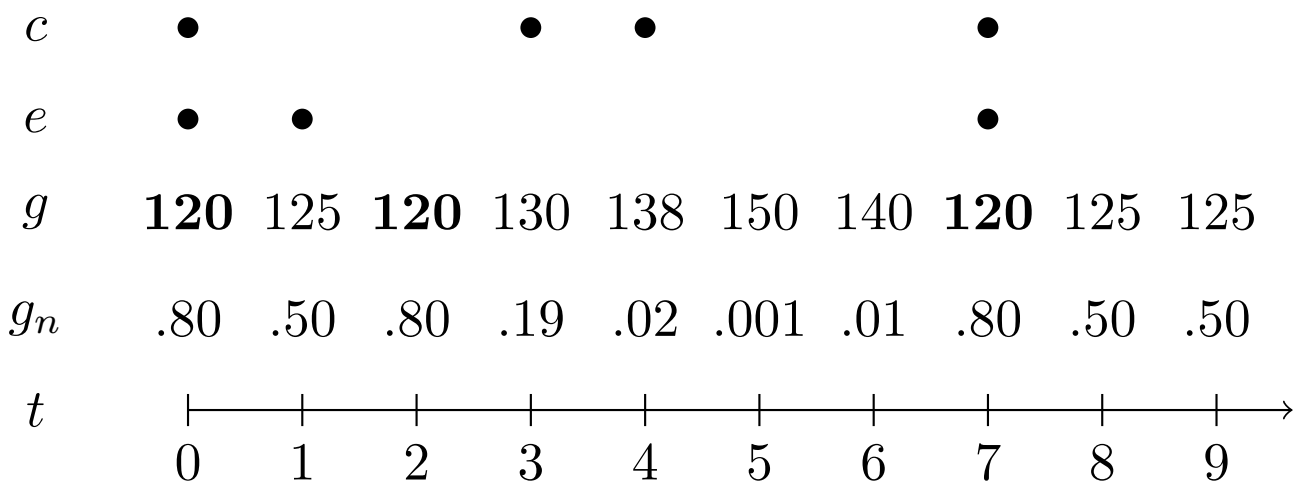
What's the effect of exercise on glucose?



Calculate: $P(g_n|c,e) - P(g_n|c,\neg e)$

Strict discretization: $= 0/2 - 0/2 = 0.$

What's the effect of exercise on glucose?



Calculate: $P(g_n|c, e) - P(g_n|c, \neg e)$

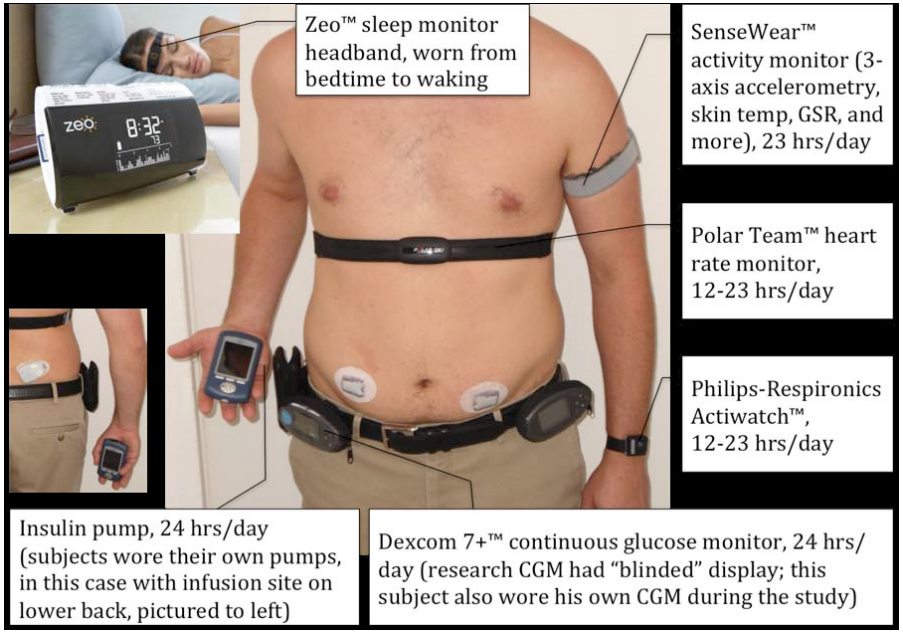
Strict discretization: $= 0/2 - 0/2 = 0.$

Probabilistic discretization:

$$= \frac{0.50 + 0.50}{2} - \frac{0.02 + 0.001}{2} = 0.49.$$

Experiment

- Cohort: 17 subjects with T1DM
- Sensor data (collected for >72 hours)
  - Glucose values
  - Insulin dosage
  - Activity
  - Sleep stage
  - Heart rate
  - Temperature



Data from N. Heintzman at UCSD

Results

$$\text{vigorous exercise} \rightsquigarrow^{\geq 15, \leq 40} \text{hyperglycemia}$$

- Effect occurs over 5-80 minutes, peaking 15-40 minutes
- Not found with strict discretization
- Supported by medical studies
  (Marliss and Vranic, 2002; Riddell and Perkins, 2006)

Conclusions

- Better representation of uncertainty
  - Captures more prior knowledge
- Uncertainty in causal inference
  - Computationally feasible
  - Increased power
  - Realistic discretization

Conclusions

Better capture uncertainty
→ Better causal inference

  - While keeping computational feasibility
- Use SOPDs in causal inference
  - Incorporate prior knowledge into (center, spread)
- Future work
  - Implement adjustment for missing, outdated data
  - Capture more of the SOPD beyond standard deviation
  - Formally prove SOPD robustness to approximation