Learning Bayesian Networks

(from completely observed data)

The Problem

- We are considering m r.v.'s, $X_1, \ldots, X_m$.
- We have a data set of n joint samples,

$$D = \left\{ \begin{array}{c} (x_1^1, x_2^1, \ldots, x_m^1), \\ (x_1^2, x_2^2, \ldots, x_m^2), \\ \vdots \\ (x_1^n, x_2^n, \ldots, x_m^n) \end{array} \right\}$$

- We want to build a Bayes net that models the joint probability distribution.

Example, from the Wisconsin breast cancer data

- Size $=$ mean tumor cell size, real or discretized into $\{$Little,Big$\}$.
- Texture $=$ mean tumor cell texture (roughness), real or discretized into $\{$Smooth,Rough$\}$.
- Recur $=$ whether or not the cancer recurred, true or false.
- Time, real or discretized $\{$1yr, 2yr, 3yr,$\ldots\}$,

$$= \begin{cases} \text{time to recurrence} & \text{if Recur} \\ \text{time free of cancer} & \text{if not Recur} \end{cases}$$

Size	Texture	Recur	Time
18.02	27.60	false	31
20.29	14.34	true	27
⋮	⋮	⋮	⋮

Size and Texture

- Consider modeling the relationship between discretized size and discretized texture.

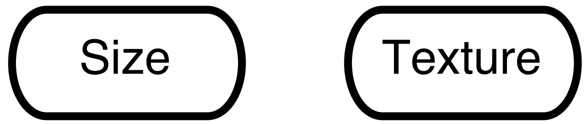
Counts	Little	Big
Smooth	62	51
Rough	39	46

Estimated probs.	Little	Big	
Smooth	0.31	0.26	0.57
Rough	0.20	0.23	0.43
	0.51	0.49	

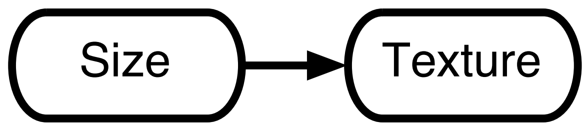
- What Bayes net structures are possible?

Three possible Bayes nets structures

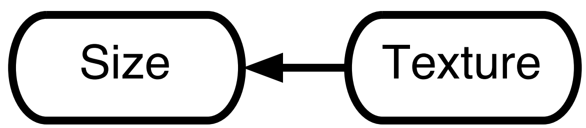
- $P(Size, Texture) = P(Size)P(Texture)$



- $P(Size, Texture) = P(Size)P(Texture|Size)$



- $P(Size, Textures) = P(Size|Texture)P(Texture)$



- Which is best? Are the variables statistically related?

Are the variables related?

	Little	Big
Smooth	62 (57.6)	51 (55.4)
Rough	39 (43.4)	46 (42.6)

- A χ^2-test gives little reason to reject independence.

$$S = 1.56, \text{p-value} > 0.1$$

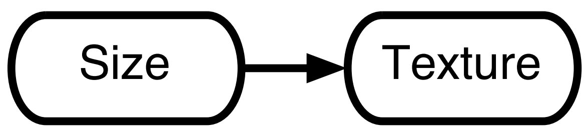
Are the variables related?

	Little	Big
Smooth	62 (57.6)	51 (55.4)
Rough	39 (43.4)	46 (42.6)

- A χ^2-test gives little reason to reject independence.

$$S = 1.56, \text{p-value} > 0.1$$

- But suppose we allow for a relationship anyway...



- How do we fill in the *parameters* of the Bayes net? P(Size) and P(Texture|Size)?

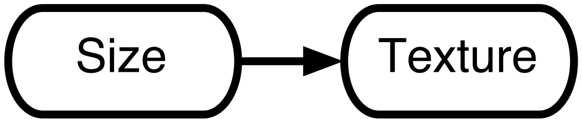
Maximum likelihood solution

- In general, suppose we have settled on a Bayes net structure for m variables, $X_1, \ldots, X_m$.
- Let $\pi_i \subset \{X_1, \ldots, X_m\}$ denote the parents of X_i.
- Let x_i^j denote the value of r.v. X_i in sample j.
- Assuming the data is i.i.d., its probability is

$$\begin{aligned}\Pi_j P(X_1 = x_1^j, \ldots, X_m = x_m^j) &= \Pi_j \Pi_i P(X_i = x_i^j | X_{\pi_i} = x_{\pi_i}^j) \\ &= \Pi_i \Pi_j P(X_i = x_i^j | X_{\pi_i} = x_{\pi_i}^j)\end{aligned}$$

- To maximize the probability of the data, we can maximize each $\Pi_j P(X_i = x_i^j | X_{\pi_i} = x_{\pi_i}^j)$ independently.
- It's like solving m independent supervised learning problems.

For the Size and Texture model

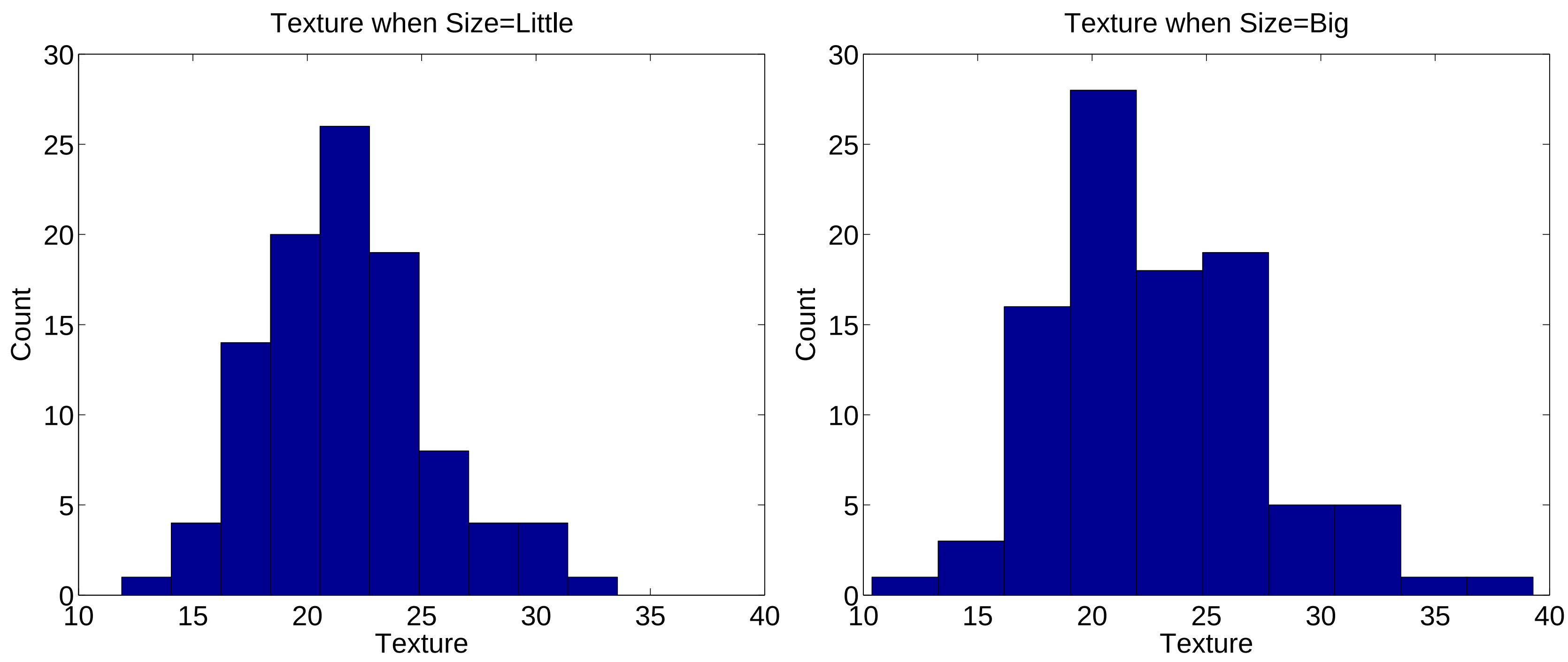


Counts	Little	Big
Smooth	62	51
Rough	39	46

- $P(Big) = \frac{51+46}{39+46+51+62} = 0.49$
- $P(Little) = \frac{62+39}{39+46+51+62} = 0.51$
- $P(Rough|Little) = \frac{39}{39+62} = 0.39$
- $P(Smooth|Little) = \frac{62}{39+62} = 0.61$
- $P(Rough|Big) = \frac{46}{51+46} = 0.47$
- $P(Smooth|Big) = \frac{51}{51+46} = 0.53$

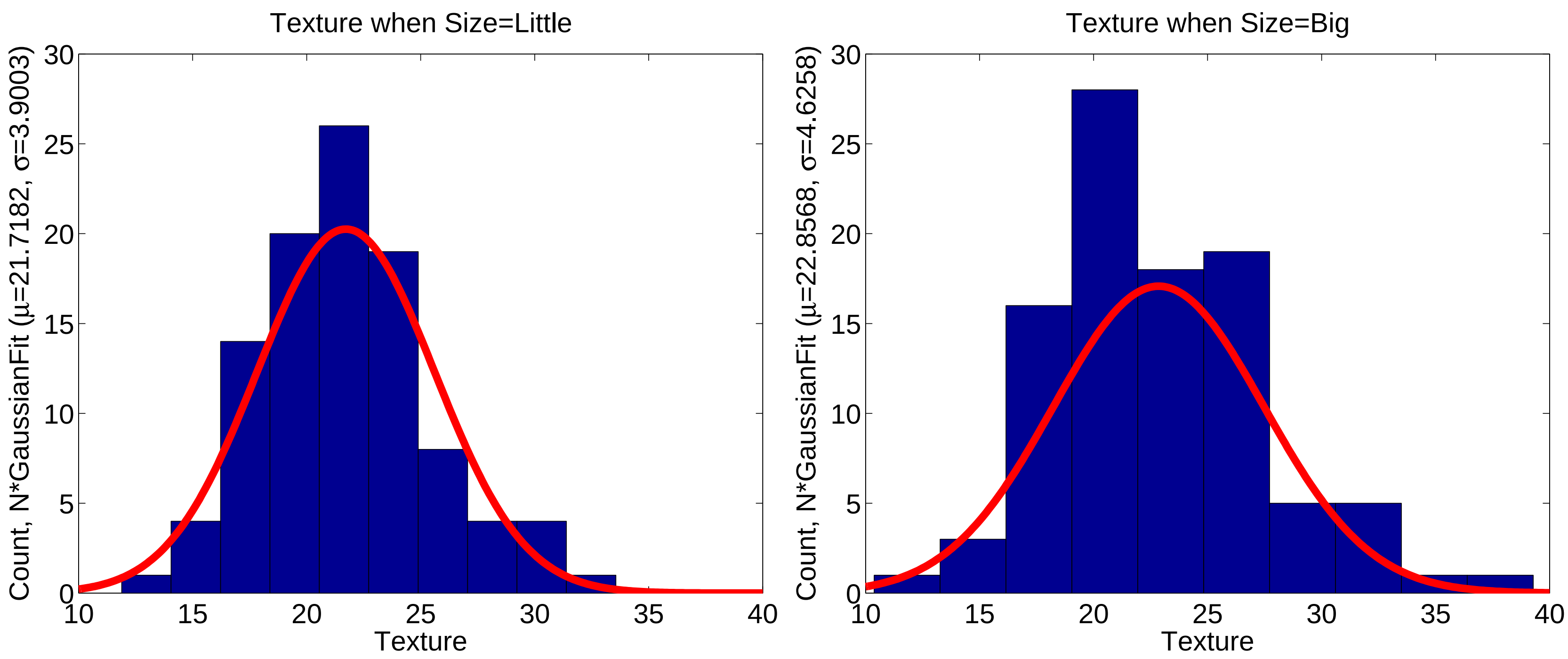
Real-valued Texture

- Suppose we use discretized size, but leave texture real-valued.
- P(Size) is fit the same.
- What do we do about P(Texture|Size)?



Real-valued Texture

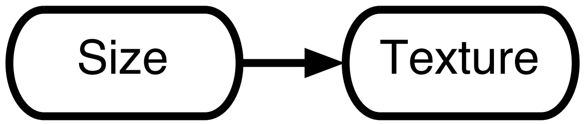
- Suppose we use discretized size, but leave texture real-valued.
- P(Size) is fit the same.
- What do we do about P(Texture|Size)? Gaussian fit?



Real-valued Texture 2

- P(Big)=0.49 (as before)
- P(Little)=0.51 (as before)
- P(Texture$|$Big)=Normal($\mu = 22.9$,$\sigma = 4.6$)
- P(Texture$|$Little)=Normal($\mu = 21.7$,$\sigma = 3.9$)

How do we make inferences?



When texture is discretized:

- What is $P(\text{Texture=Rough} \mid \text{Size=Big})$?
- What is $P(\text{Size=Big} \mid \text{Texture=Rough})$?

When texture is real-valued:

- What is $P(\text{Texture=17} \mid \text{Size=Big})$?
- What is $P(\text{Size=Big} \mid \text{Texture=17})$?

How can we model all four r.v.'s?

- Size $=$ mean tumor cell size, real or discretized into $\{$Little,Big$\}$.
- Texture $=$ mean tumor cell texture (roughness), real or discretized into $\{$Smooth,Rough$\}$.
- Recur $=$ whether or not the cancer recurred, true or false.
- Time, real or discretized $\{$1yr, 2yr, 3yr,$\ldots\}$,

$$= \begin{cases} \text{time to recurrence} & \text{if Recur} \\ \text{time free of cancer} & \text{if not Recur} \end{cases}$$

Summary

- Bayes net structures can be chosen based on:
  - The meaning of the r.v.'s
  - Statistical tests
  - Convenience
- Once the structure is given, the parameters (conditional p.d.f.'s for each r.v.) can be optimized independently according to maximum likelihood.