Unit 6: Bayesian Statistics

3. Graphical Models

4/27/2020

Recap from last time

1. Likelihood ratios give us a way to compare models
   (the step function is approximating this)

2. Bayesian inference naturally encodes a preference for simpler models through posterior averaging

3. We can infer the values of unknown parameters in a way that reflects both the data and our prior beliefs

Key ideas

1. Graphical models are a way of visually depicting the statistical relationships among variables in a model

2. You can use graphical models to investigate causal relationships including Bayesian explaining away

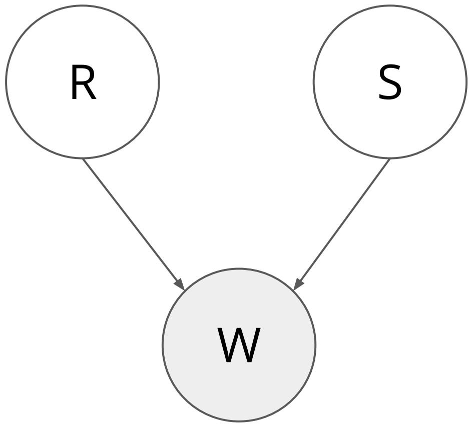
3. This same framework can depict the kinds of models that we have worked with in the frequentist framework. This can be helpful for thinking about the data generation process

Graphical Models

Graphical models are a visual notation for expressing the probabilistic relationships among a set of variables.

Components:

1. **Vertices** that represent the variables
2. **Edges** that represent statistical dependencies between the vertices
3. A set of **probability distributions** that describe these dependencies

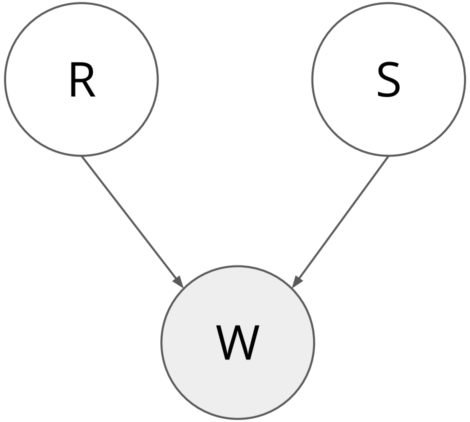


Latent and Observed Variables

Vertices represent two kinds of variables:

1. **Observed variables** (filled circles) are variables whose values we see directly.
2. **Latent variables** (empty circles) are variables that we do not see, but that explain the process that generated the observed variables.

Typically, we want to infer the values of the latent variables from the observed variables in our data



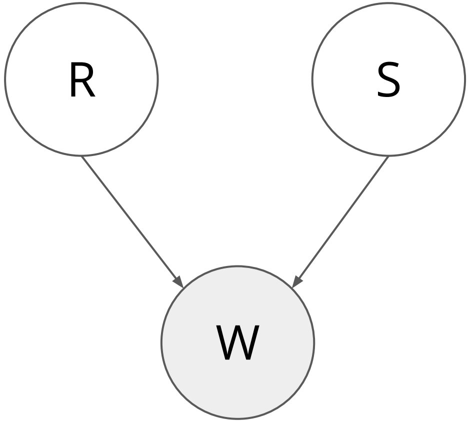
A graphical model for wet grass

This simple model describes how grass might get wet

W denotes whether grass is wet or dry. We is an observed variables because we get to see it

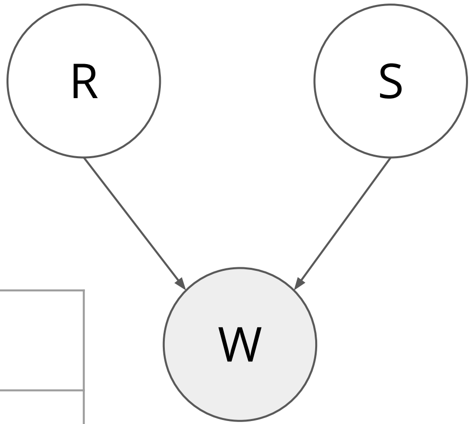
R (rain) and **S** (sprinklers) are potential causes of wet grass. They are latent because we don’t get to observe them

Because there is no arrow between **R** and **S**, we know that they are independent



Filling out the graphical model

P(R) = .4



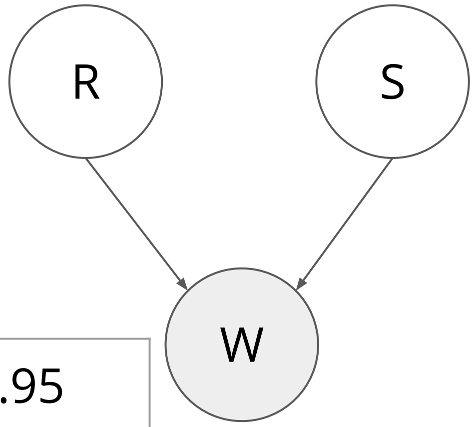
P(S) = .2

P(W|S & R) = .95
P(W|S & ~R) = .9
P(W|~S & R) = .9
P(W|~S & ~R) = .1

Using the model to reason forward

Suppose we *know* the sprinklers turned on. What is the probability that the grass is wet?

P(R) = .4



P(S) = .2

$$P(W|S) = P(W\,|\,S\,\&\,R)P(R) + P(W\,|\,S\,\&\,{\sim}R)P({\sim}R)$$

P(W|S & R)	= .95
P(W|S & ~R)	= .9
P(W|~S & R)	= .9
P(W|~S & ~R)	= .1

P(W|S) = .95 * .4 + .90 * .6
= .92

Using the model to reason backward

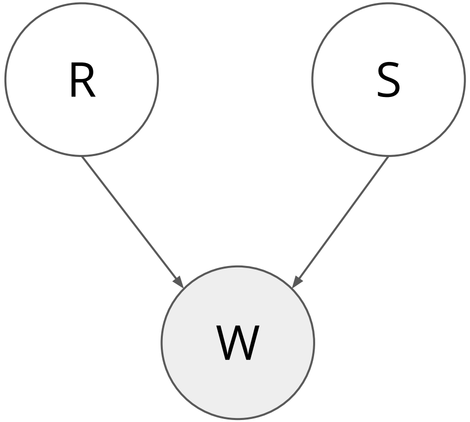
Suppose we *know* the grass is wet.
What is the probability the sprinklers turned on?

$$P(S|W) = \frac{P(W|S)P(S)}{P(W)} = \frac{.92 \cdot .2}{.52} \sim .35$$

$$\begin{aligned} P(W) = {} & P(W \mid S \,\&\, R)P(S)P(R) \\ & + P(W \mid S \,\&{\sim}R)P(S)P({\sim}R) \\ & + P(W \mid {\sim}S \,\&R)P({\sim}S)P(R) \\ & + P(W \mid {\sim}S \,\&{\sim}R)P({\sim}S)P({\sim}R) \end{aligned}$$

$$= .95 \cdot .2 \cdot .4 + .9 \cdot .2 \cdot .6 + .9 \cdot .8 \cdot .4 + .1 \cdot .8 \cdot .6 = .52$$

P(R) = .4



P(S) = .2

P(W | S & R)	= .95
P(W | S & ~R)	= .9
P(W | ~S & R)	= .9
P(W | ~S & ~R)	= .1

Using the model to diagnose hidden causes

Suppose we *know* the grass is wet *and that it rained*.
What is the probability the sprinklers turned on?

$$\mathrm{P}(S|W \,\&\, R) = \frac{P(W \,\&\, S \,\&\, R)}{P(W \,\&\, R)}$$

$$= \frac{P(W|S \,\&\, R)P(S \,\&\, R)}{P(W \,\&\, R)}$$

$$= \frac{P(W|S \,\&\, R)P(S)\cancel{P(R)}}{P(W|R)\cancel{P(R)}}$$

$$= \frac{P(W|S \,\&\, R)P(S)}{P(W|S \,\&\, R)P(S) + P\left(W \middle| {\sim}S \,\&\, R\right)P({\sim}S)}$$

$$= \frac{.95 \cdot .2}{.95 \cdot .2 \ + \ .9 \cdot .8} = .21$$

Explaining away

We just discovered something interesting!

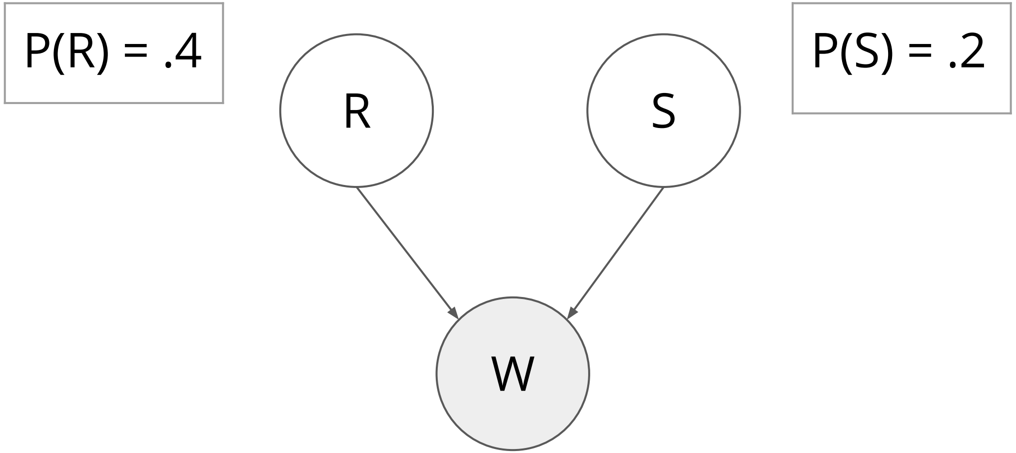
$$P(S|W) = .35$$

$$P(S|W \,\&\, R) = .21$$

The sprinklers and the rain are independent of each-other.

But they are conditionally-dependent on each other through the wetness of grass

Rain **explains away** sprinklers as a cause of wet grass

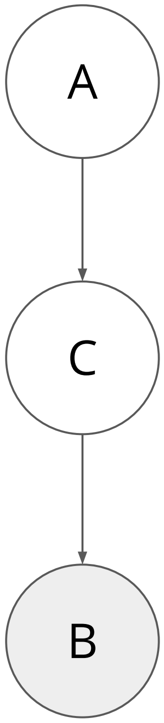


Conditional independence

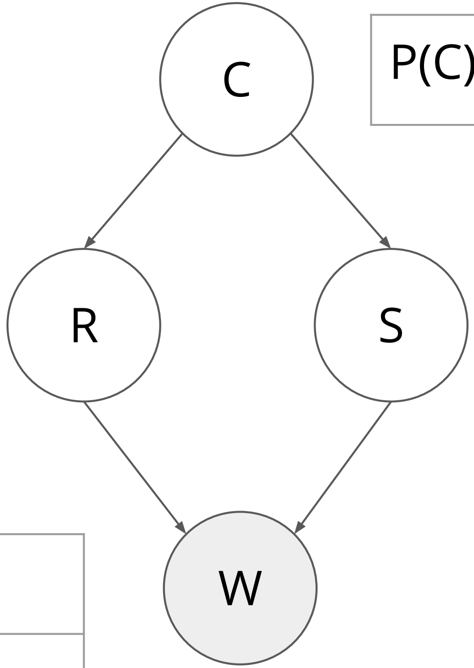
Events A and B are **independent** if P(A & B) = P(A)P(B)

Events A and B are **conditionally independent** given event C
if P(A | B & C) = P(A | B)

In a graphical model, grand-children of a vertex are independent of their grandparents given their children



Conditional independence



$P(C) = .5$

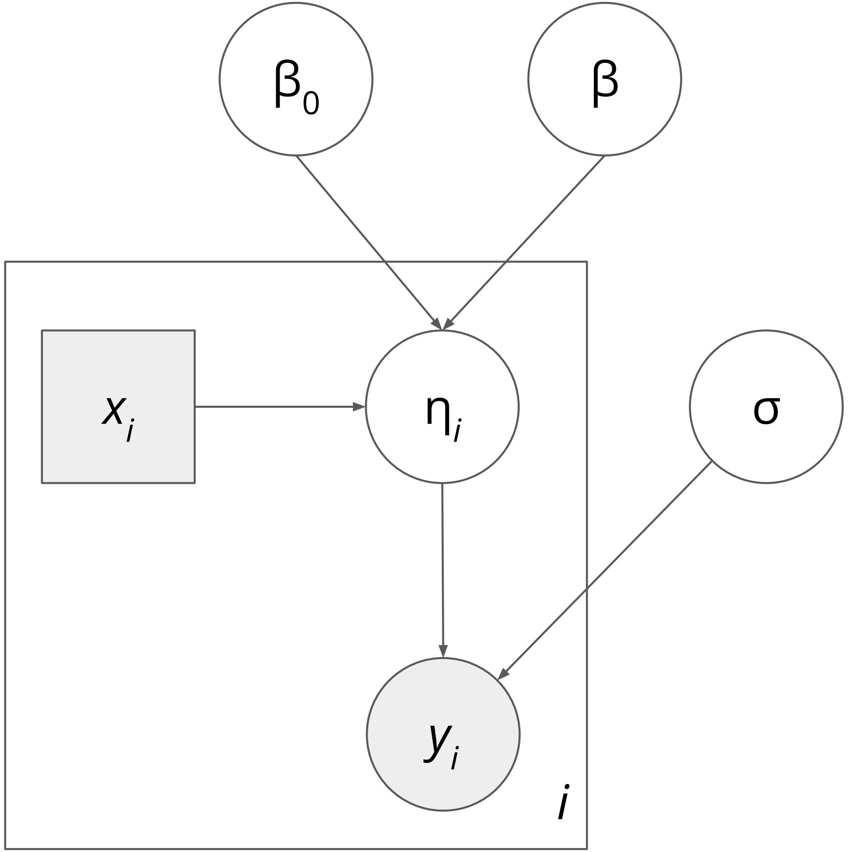
$P(R \mid C) = .8$
$P(R \mid {\sim}C) = .1$

$P(S \mid C) = .1$
$P(S \mid {\sim}C) = .5$

$P(W \mid S \,\&\, R) = .95$
$P(W \mid S \,\&\, {\sim}R) = .9$
$P(W \mid {\sim}S \,\&\, R) = .9$
$P(W \mid {\sim}S \,\&\, {\sim}R) = .1$

Bayesian Linear Regression

We want to model the relationship between independent variables *x* and dependent variables *y*.



Frequentist linear regression

The regression models you've already seen are a special case of the Generalized Linear Model (GLM)

1. A probability distribution describing the outcome:

$$y_i = \text{Normal}(p_i, \sigma^2)$$

2. A linear model:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

3. A link function that relates the linear model to the parameter of the outcome distribution

$$p = \eta$$

Bayesian Linear Regression

We want to model the relationship between independent variables *x* and dependent variables *y*.

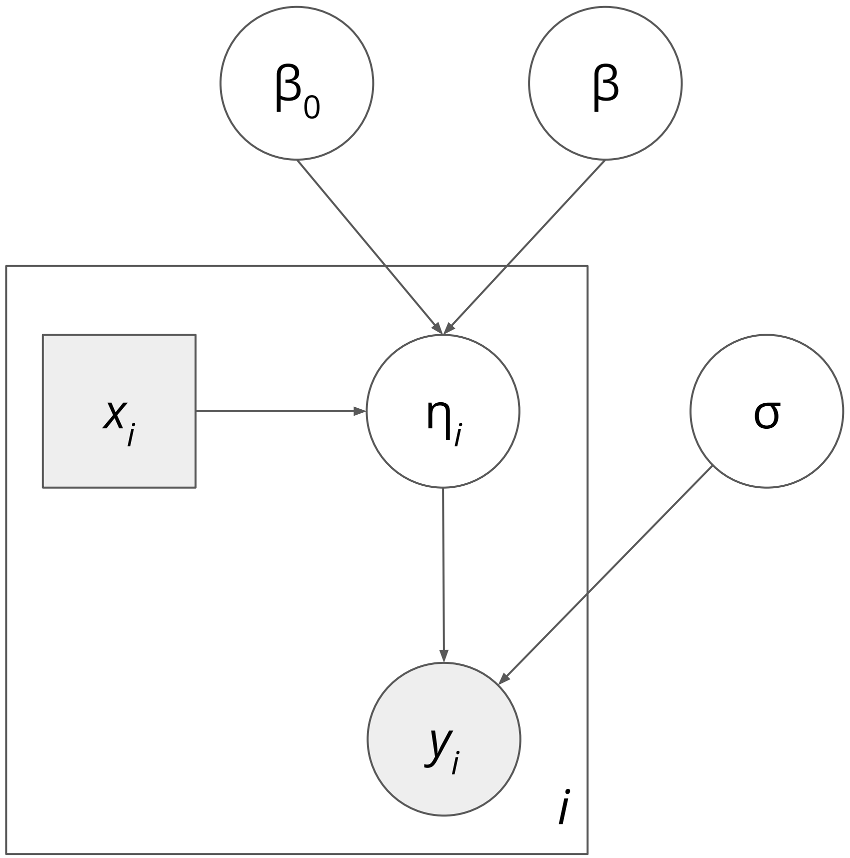
Likelihood

$$\eta_i = \beta_0 + \beta \cdot x_i$$

$$y_i \sim \mathrm{Normal}(\eta_i, \sigma)$$

Prior

$$\sigma \propto \frac{1}{\sigma^2} \qquad \beta_0, \beta \propto 1$$



Frequentist and Bayesian Models

Frequentist Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4045.33	286.21	14.1	<2e-16 ***
carat	7765.14	14.01	554.3	<2e-16 ***
depth	-102.17	4.64	-22.0	<2e-16 ***

Multiple R-squared: 0.851,   Adjusted R-squared: 0.851
F-statistic: 1.54e+05 on 2 and 53937 DF, p-value: <2e-16

Bayesian Model

	mean	sd	10%	50%	90%
(Intercept)	4045.4	290.9	3672.5	4041.5	4422.6
carat	7765.1	13.9	7746.9	7764.8	7783.1
depth	-102.2	4.7	-108.3	-102.1	-96.1
sigma	1541.9	4.5	1536.1	1541.8	1547.7

Key ideas

1. Graphical models are a way of visually depicting the statistical relationships among variables in a model

2. You can use graphical models to investigate causal relationships including Bayesian explaining away

3. This same framework can depict the kinds of models that we have worked with in the frequentist framework. This can be helpful for thinking about the data generation process