More on interactions

ISTA 410 / INFO 510 - Bayesian Modeling and Inference

University of Arizona School of Information

September 22, 2021

Outline

Last time:

- Causal DAGs; unobserved variables in DAGs
- Intro to interactions

Today: more interactions

Modeling interactions

Interactions

Interaction effects:

- effect of one predictor is *conditional* on another
    - Effect of *water* on *plant growth* is conditional on *sunlight*
    - Effect of *gene* on *survival* is conditional on *environment*
    - Effect of *total traffic* on *bike traffic* is conditional on *bike lane*
- Interactions appear frequently in real systems

Interactions in DAGs

Here is what an interaction looks like in a DAG:

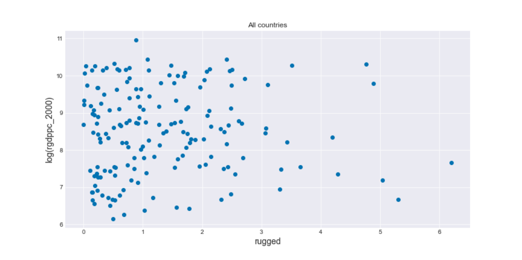
Any time two variables influence a third leaves the possibility for interaction!

Terrain "ruggedness" and economy

(Example from Statistical Rethinking Ch7.) What is the relationship between the geographic terrain in a nation and its economy?

Data: observations on many countries

- Outcome: log GDP (as of 2000, when data was collected)
- Predictor: terrain "ruggedness" index

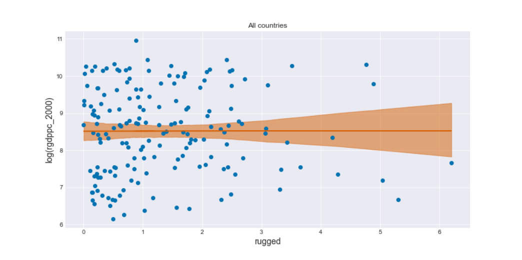


Terrain "ruggedness" and economy

(Example from Statistical Rethinking Ch7.) What is the relationship between the geographic terrain in a nation and its economy?

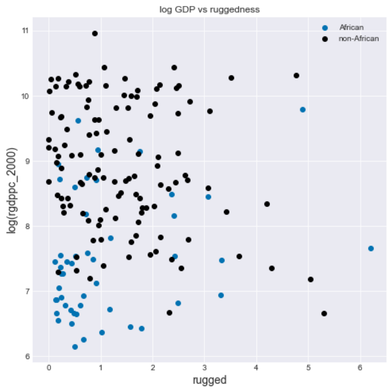
Data: observations on many countries

- Outcome: log GDP (as of 2000, when data was collected)
- Predictor: terrain "ruggedness" index



Terrain ”ruggedness” and economy

Closer examination of the data reveals an interesting phenomenon: the relationship is different for countries in Africa.



Terrain ”ruggedness” and economy

Conditionality:

- The effect of ruggedness on modern economy is *conditional* on continent
- African nations respond differently to ruggedness than non-African nations

Want to incorporate this effect into a model; ideally, a *single* model

- Pooling data gives better estimates of continent-independent parameters

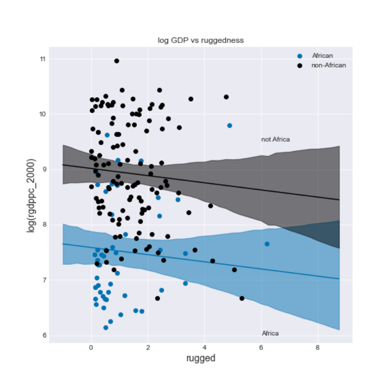
A simple approach that won't work

A simple approach that's not quite good enough: add an indicator variable for African countries, and do a bivariate regression:

$$
\begin{aligned}
\log GDP &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_R R_i + \beta_A A_i \\
\beta_R &\sim \mathrm{Normal}(0, 1) \\
\beta_A &\sim \mathrm{Normal}(0, 1) \\
\alpha_j &\sim \mathrm{Normal}(9, 3) \\
\sigma &\sim \mathrm{HalfCauchy}(5)
\end{aligned}
$$

A simple approach that won't work

Problem:



Allows for a shift, but not a change in slopes.

Can see this also with the fox problem from earlier.

Allowing interactions

To add interactions:

$$
\begin{aligned}
\log GDP &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_R R_i + \beta_A A_i + \beta_{AR} A_i R_i \\
\beta_R &\sim \mathrm{Normal}(0, 1) \\
\beta_A &\sim \mathrm{Normal}(0, 1) \\
\beta_{AR} &\sim \mathrm{Normal}(0, 1) \\
\alpha_j &\sim \mathrm{Normal}(9, 3) \\
\sigma &\sim \mathrm{HalfCauchy}(5)
\end{aligned}
$$

So we have a third slope, for the *product* of R and A.

Why is this the approach?

Where this comes from: just model the slope β_R as being itself a linear function of A.

$$
\begin{aligned}
\log GDP &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \gamma_i R_i + \beta_A A_i \\
\gamma_i &= \beta_R + \beta_{AR} A_i \\
\beta_R &\sim \mathrm{Normal}(0, 1) \\
\beta_A &\sim \mathrm{Normal}(0, 1) \\
\beta_{AR} &\sim \mathrm{Normal}(0, 1) \\
\alpha_j &\sim \mathrm{Normal}(9, 3) \\
\sigma &\sim \mathrm{HalfCauchy}(5)
\end{aligned}
$$

Why is this the approach?

Plug a linear equation into another linear equation:

$$\mu_i = \alpha + \gamma_i R_i + \beta_A A_i$$

$$\gamma_i = \beta_R + \beta_{AR} A_i$$

Why is this the approach?

Plug a linear equation into another linear equation:

$$\mu_i = \alpha + \gamma_i R_i + \beta_A A_i$$

$$\gamma_i = \beta_R + \beta_{AR} A_i$$

$$\mu_i = \alpha + (\beta_R + \beta_{AR} A_i) R_i + \beta_A A_i$$

Code for the interaction model

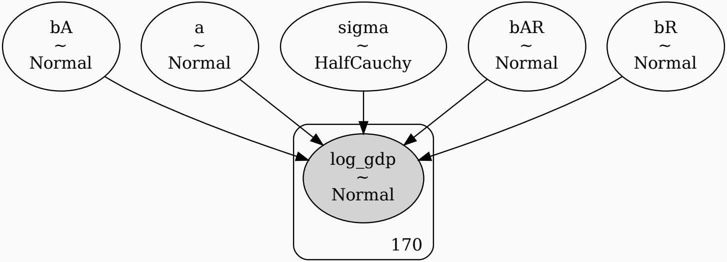
```python
with pm.Model() as model_product:
    # Priors
    a = pm.Normal('a', mu=9, sd=3)
    bR = pm.Normal('bR', mu=0, sd=1)
    bA = pm.Normal('bA', mu=0, sd=1)
    bAR = pm.Normal('bAR', mu=0, sd=1)
    sigma = pm.HalfCauchy('sigma', 5)

    # Model equations -- could write these differently
    gamma = bR + bAR * data.cont_africa
    mu = a + gamma * data.rugged + bA * data.cont_africa

    # Likelihood
    log_gdp = pm.Normal('log_gdp',
                        mu,
                        sigma,
                        observed=np.log(data.rgdppc_2000))

    # Inference
    qp_product = quap()
```

The result

Result from the interaction model:

Here, we can see the different slope. Ruggedness has a positive association with GDP for African nations, negative for others.

Plate diagram for the product version



The Africa interaction as two independent slopes

Alternative formulation of the model:

$$\begin{aligned}
\log GDP &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\mathrm{CONT}[i]} + \beta_{CONT[i]} R_i \\
\beta_j &\sim \mathrm{Normal}(0, 1) \qquad (j \in \{0, 1\}) \\
\alpha_j &\sim \mathrm{Normal}(9, 3) \qquad (j \in \{0, 1\}) \\
\sigma &\sim \mathrm{HalfCauchy}(5)
\end{aligned}$$

- *index variable*
- α, β are now length-2 vectors – one intercept and one slope for each
- allows the model to simply select separate slopes independently

Aside: categorical variables

Categorical variables in regression

Ways to handle a categorical variable c:

- indicator variable
- if c is binary, assign value 0 to one category, 1 to another

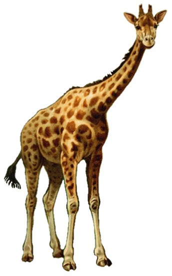
$$\mu_j = \alpha + \beta_c c$$

Potential problem: more uncertainty in one category than another

Example: giraffe height

Simple example: modeling giraffe height stratified by sex. Assign $s = 0$ for female, $s = 1$ for male.

$$\begin{aligned} h &\sim \mathrm{Normal}(\mu, \sigma) \\ \mu &= \alpha + \beta s \\ \alpha &\sim \mathrm{Normal}(5, 1.5) \\ \beta &\sim \mathrm{Normal}(0, 0.4) \end{aligned}$$



- $\mathrm{Var}\,\mu$ for males $= \mathrm{Var}\,\alpha + \mathrm{Var}\,\beta$
- $\mathrm{Var}\,\mu$ for females $= \mathrm{Var}\,\alpha$

Alternatives: one-hot encoding or index variables

One alternative: one-hot encoding

- create an indicator variable for every category

$$\mu = \beta_f f + \beta_m m$$

Why drop α? Not enough constraints:

- Suppose female giraffes average 4.8 m, males 5.1; then which is correct?

$$\mu = 5 + 0.1m - 0.2f$$

$$\mu = 4 + 1.1m + 0.8f$$

- Model is not *identifiable* – different parameters produce the same probability distribution

Alternatives: one-hot encoding or index variables

Another: index variables

- create a vector of intercepts and use the value of the categorical variable to index out the right one

$$\mu = \beta_s \qquad s \in \{0, 1\}$$

  Requires encoding values of the variable as ordinal values $\{0, 1, 2, \ldots, n\}$

This is what we did with the Africa ruggedness example:

- Africa indicator variable has values 0, 1
- Use these as an index to select the correct parameter from the vector

Code for the index variable model

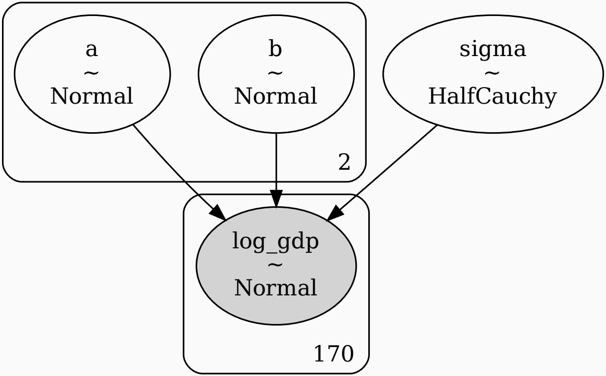
```
with pm.Model() as model_index:
    # Priors
    a = pm.Normal('a', 9, 3, shape=2)
    b = pm.Normal('b', 0, 1, shape=2)
    sigma = pm.HalfCauchy('sigma', 5)

    # Model equation
    mu = a[dd.cont_africa] + b[dd.cont_africa] * dd.rugged

    # Likelihood
    log_gdp = pm.Normal('log_gdp',
                        mu,
                        sigma,
                        observed=np.log(dd['rgdppc_2000']))

    # Inference
    qp_index = quap()
```

Mathematical form:

$$
\begin{aligned}
\log GDP &\sim \mathrm{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\mathrm{CONT}[i]} + \beta_{CONT[i]} R_i \\
\beta_j &\sim \mathrm{Normal}(0, 1) \qquad (j \in \{0, 1\}) \\
\alpha_j &\sim \mathrm{Normal}(9, 3) \qquad (j \in \{0, 1\}) \\
\sigma &\sim \mathrm{HalfCauchy}(5)
\end{aligned}
$$

- Note shape parameter in setting up variables
- indexing with a data frame column – mu has an entry for each row in the data frame
- which component of a or b contributes to mu depends on the value of cont_africa in the corresponding row

Plate diagram for the index version



Comparing parameters

- Parameters from the product-interaction model:

	mean	sd	hdi_3%	hdi_97%
a	9.183	0.136	8.927	9.439
bR	-0.184	0.076	-0.326	-0.042
bA	-1.846	0.218	-2.256	-1.435
bAR	0.348	0.127	0.108	0.588
sigma	0.933	0.051	0.838	1.028

- Parameters from the index-variable model:

	mean	sd	hdi_3%	hdi_97%
a[0]	9.221	0.138	8.962	9.480
a[1]	7.283	0.176	6.952	7.615
b[0]	-0.201	0.076	-0.345	-0.058
b[1]	0.186	0.105	-0.010	0.383
sigma	0.933	0.051	0.837	1.028

Higher order interactions

Higher order interactions

No interactions:

$$\mu_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i}$$

Higher order interactions

No interactions:

$$\mu_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i}$$

Pairwise interactions:

$$\begin{aligned}\mu_i = \alpha &+ \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} \\ &+ \beta_{12} x_{1,i} x_{2,i} + \beta_{13} x_{1,i} x_{3,i} + \beta_{23} x_{2,i} x_{3,i}\end{aligned}$$

Higher order interactions

No interactions:

$$\mu_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i}$$

Pairwise interactions:

$$\begin{aligned}\mu_i = \alpha &+ \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} \\ &+ \beta_{12} x_{1,i} x_{2,i} + \beta_{13} x_{1,i} x_{3,i} + \beta_{23} x_{2,i} x_{3,i}\end{aligned}$$

Higher-order interactions:

$$\begin{aligned}\mu_i = \alpha &+ \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} \\ &+ \beta_{12} x_{1,i} x_{2,i} + \beta_{13} x_{1,i} x_{3,i} + \beta_{23} x_{2,i} x_{3,i} \\ &+ \beta_{123} x_{1,i} x_{2,i} x_{3,i}\end{aligned}$$

The ”Judgement of Princeton”

The Judgement of Princeton

- 9 judges, 20 wines
- Wines split between red and white, NJ or France
- Judges split between American or French/Belgium

Predictors:

- Wine color: red or white
- Wine origin: NJ or France
- Judge nationality: US or EU

Interactions

Potential for interactions between all predictors:

- Interaction between origin and judge: judge bias.
  Judge bias might depend upon color.
- Interaction between color and judge: taste preference.
  Taste preference might depend upon origin.
- Interaction between origin and color: relative advantage.
  Advantage might depend upon judge.

Comparing predictions

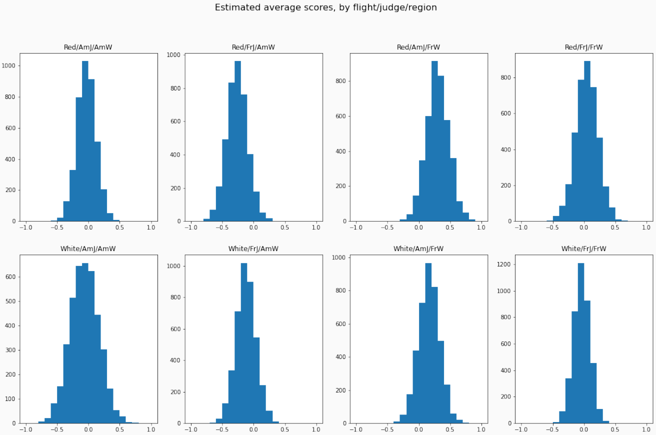
- interpreting interaction coefficients directly: hard!
- here, predictors are discrete:
  - draw a sample from the posterior distribution
  - each sample of slopes represents a prediction for each combination of categories
  - compare predictions for pairs of combination

Version 1

First version, with indicator variables for region, color, and judge

```
with pm.Model() as interaction:
    # Priors
    # omitted for space

    # Model equation
    mu = (alpha + bC * wines['flight'] + bJ * wines['judge.amer'] + bR * wines['wine.amer']
          + bJR * wines['judge.amer'] * wines['wine.amer']
          + bCJ * wines['judge.amer'] * wines['flight']
          + bRC * wines['flight'] * wines['wine.amer'])

    # Likelihood
    score = pm.Normal('score', mu=mu, sigma=sigma, observed=wines['score'])
```

Results from version 1



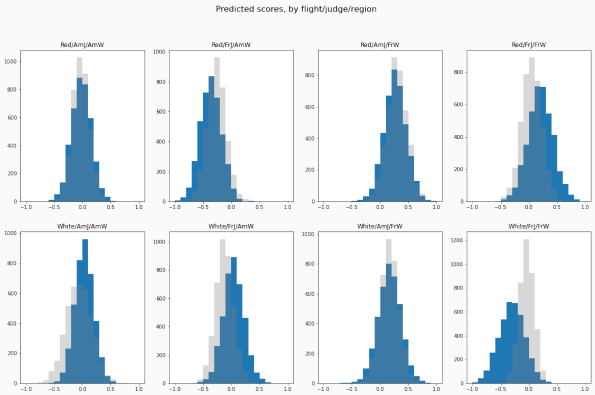
Version 2

Second version, with a separate “slope” for each combination of predictors

```
with pm.Model() as interaction:
    # Priors
    b = pm.Normal('b', 0, 0.5, shape = (2, 2, 2))
    sigma = pm.Exponential('sigma', 1)
    alpha = pm.Normal('alpha', 0, 0.2)
    # Model equation
    mu = b[wines['flight'], wines['judge.amer'], wines['wine.amer']]

    # Likelihood
    score = pm.Normal('score', mu=mu, sigma=sigma, observed=wines['score'])
```

- $2 \times 2 \times 2$ array of parameters!
- 3-dimensional indexing – easier organization here

Results from version 2



Notice difference in lower right – why?

- White wine, French judge, French wine all coded as 0

Continuous interactions

Continuous interactions:

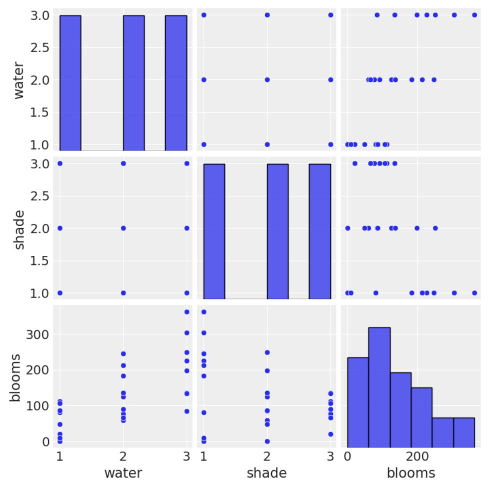
- Handled similarly to discrete interactions:
  - Interaction term between x_1, x_2:

$$\beta_{1,2} x_1 x_2$$

- Interpretation can be trickier than with discrete variables – no comparison across categories
- Example from book:
  - Tulip blooms as a function of water and light

Continuous interactions

- Tulips – flowers cultivated for blooms
- Plants provided with water and sunlight/shade
- Interested in interaction of water and sunlight



Continuous interactions

Non-interaction model:

$$b_i \sim \mathrm{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_w(w_i - \bar{w}) + \beta_s(s_i - \bar{s})$$

- independent linear effects for light and shade
- realistically: water does not help if plants don't get sunlight

Continuous interactions

Non-interaction model:

$$b_i \sim \mathrm{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_w(w_i - \bar{w}) + \beta_s(s_i - \bar{s})$$

- independent linear effects for light and shade
- realistically: water does not help if plants don't get sunlight

Interactions:

$$B_i \sim \mathrm{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_w(w_i - \bar{w}) + \beta_s(s_i - \bar{s}) + \beta_{w,s}(w_i - \bar{w})(s_i - \bar{s})$$

Linear model for the slope

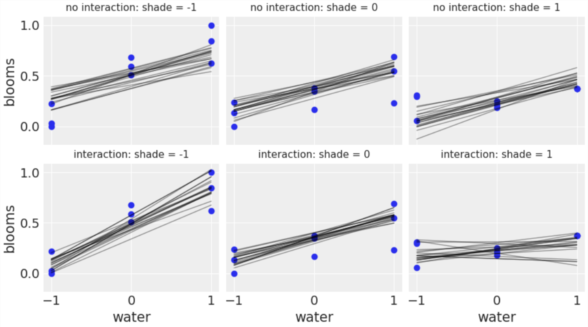
This is still just the same as using a linear model for the second slope:

$$\begin{aligned}\mu_i &= \alpha + \gamma_{w,i}(w_i - \bar{w}) + \beta_s(s_i - \bar{s}) \\ \gamma_{w,i} &= \beta_w + \beta_{w,s}(s_i - \bar{s})\end{aligned}$$

Substituting:

$$\mu_i = \alpha + (\beta_w + \beta_{w,s}(s_i - \bar{s}))(w_i - \bar{w}) + \beta_s(s_i - \bar{s})$$

Tulips



- Slopes small for $s = 1$
- Confirms our intuition that it is (water + light) that stimulates growth

Difficulty in estimating interactions

- Interaction effects can be tricky to estimate
  - Commonly smaller effects
  - Commonly noisier effects
- Rule of thumb (from Gelman et al.):
  - Need 4x the sample size to estimate an interaction effect the same size as baseline effects
  - Need 16x the sample size to estimate an interaction effect half the size as baseline effects

Summary

Today:

- Interactions

Next week:

- Information theory
- Model comparison using:
  - Information criteria
  - Approximate cross-validation