Intro to Measurement Models

Nathan Danneman

October 23, 2020

A Brief Intro to Measurement Models

This talk aims to introduce you to the theory and practice of *measurement models*.

Overview:

1. Theory: what is a measurement model
2. Bayesian estimation (just a smidge)
3. Practice: estimation in R+JAGS

Presumes:

- Moderate stats knowledge
- Working knowledge of R

What is a Measurement Model?

Most broadly, measurement models are concerned with estimating latent variables from observed variables theorized to cause or be caused by those latent variables.

They are useful for quantifying concepts whose causes or consequences are observable/measurable, but that are not themselves measurable directly.

Example: Math aptitude

- Causes:
  - math classes
- Consequences:
  - test scores

More Examples

This class of models is really common in social sciences, especially psychology. The encompassing family of methods is called Structural Equation Modeling.

Some latent variables:

- democracy
- intelligence
- perceived health
- ideology

Your own examples?

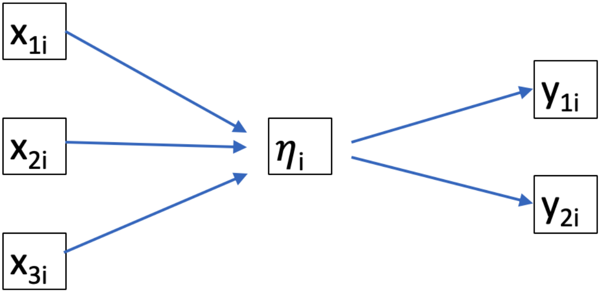
What's Needed

For our purposes, I'll be making some simplifying assumptions:

1. You've got exactly one latent variable.
   Moderately important, and hard to validate empirically
2. You have more than one other variable.
   Pretty trivial
3. You know whether your indicators cause, or are caused by, your latent variable.
   Outside of this, you need full-on SEM, which is out of scope for this talk

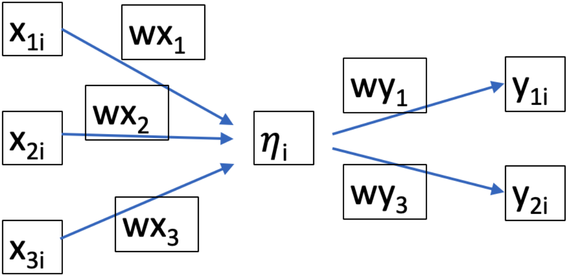
Visual: Theory

Suppose you think variables $x_{1:3}$ cause latent variable η, which causes observed outcomes $y_{1:2}$ in each of $i \in N$ observations.



Visual: Estimates

The data (inputs) are X (Nx3) and Y (Nx2).
The estimates are W (5) and η (Nx1)



Note: We may be centrally interested in η and/or w, for inference over these observations, theory-testing, or future prediction.

Bayesian Estimation, in Brief

In broad terms, Bayesian estimation involves multiplying the likelihood of seeing some data, $P(X|\theta)$ by a prior likelihood of the parameter, $P(\theta)$.

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

Solving for $P(\theta|X)$ analytically is often intractable in real cases. Also, we can re-normalize afterwards, so can frequently ignore $P(X)$. Instead, it's frequently possible to *sample* from the *marginal* distributions of parameters.

Sampling

Markov Chain Monte Carlo (MCMC) is a common method for sampling from marginal distributions.

Idea: Given the data and estimates of all other values, sample from the distribution of values for some parameter. Continue this until you have lots of samples for all the parameters, one at a time.

In practice, there are lots of better ways to do this, but that's what JAGS/BUGS does under the hood.

Estimating a Bayesian Model with JAGS

JAGS: Just Another Gibbs Sampler
Handles figuring out the distribution(s) to sample from, and does the sampling. MAGIC!

`http://mcmc-jags.sourceforge.net/`

Workflow:

1. Provide data
2. Specify likelihoods (as distributions) and relationships between data and parameters
3. Specify priors for parameters
4. Retrieve samples from posterior distribution(s) of parameter(s)
5. Evaluate the samples, check for convergence

To The Code!

```r
# Suppose we have some iid measurements.
# What is the mean and sd of the distribution from which this data was drawn?
x <- c(3, 5, 9, 9, 12, 12)
hist(x)

# The sample mean is the maximum likelihood estimator
ml_mean <- mean(x)
abline(v=ml_mean, col="red", lwd=3)

# What if you think this data was drawn from a distribution with a higher mean?
# But you aren't TOO sure?
# Suppose your prior is normally distributed, mean=10, sd=3

```

datamachines