Generalized Measurement Models: Modeling Observed Data

Lecture 4b

https://jonathantemplin.com/bayesian-psychometric-modeling-fall-2022/

Today's Lecture Objectives

1. Show different modeling specifications for different types of item response data

2. Show how parameterization differs for standardized latent variables vs. marker item scale identification

Example Data: Conspiracy Theories

Today's example is from a bootstrap resample of 177 undergraduate students at a large state university in the Midwest. The survey was a measure of 10 questions about their beliefs in various conspiracy theories that were being passed around the internet in the early 2010s. Additionally, gender was included in the survey. All items responses were on a 5-point Likert scale with:

1. Strongly Disagree

2. Disagree

3. Neither Agree or Disagree

4. Agree

5. Strongly Agree

PLEASE NOTE, THE PURPOSE OF THIS SURVEY WAS TO STUDY INDIVIDUAL BELIEFS REGARDING CONSPIRACIES. THE QUESTIONS CAN PROVOKE SOME STRONG EMOTIONS GIVEN THE WORLD WE LIVE IN CURRENTLY. ALL QUESTIONS WERE APPROVED BY UNIVERSITY IRB PRIOR TO THEIR USE.

Our purpose in using this instrument is to provide a context that we all may find relevant as many of these conspiracy theories are still prevalent today.
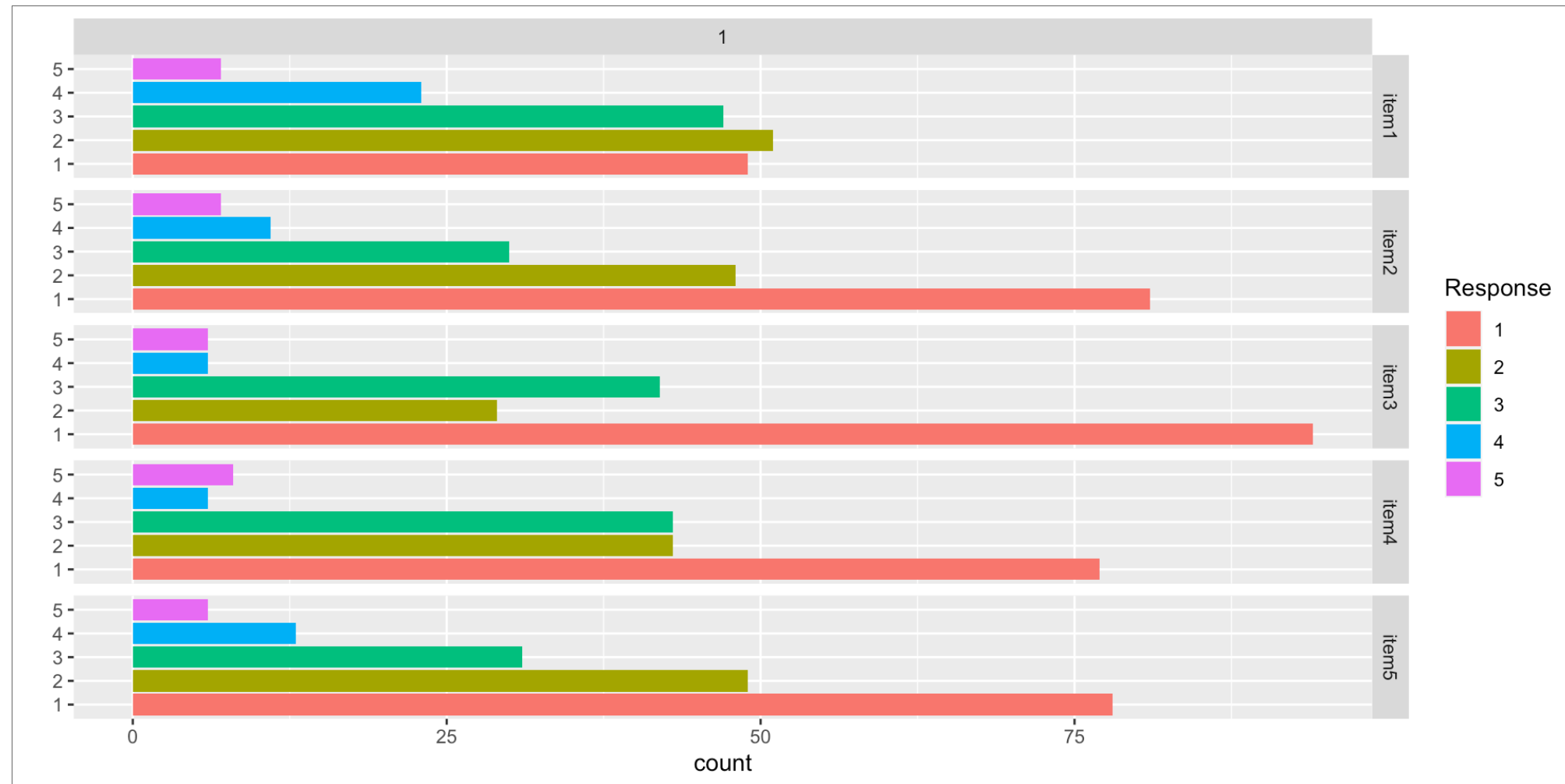
Conspiracy Theory Questions 1-5

Questions:

1. The U.S. invasion of Iraq was not part of a campaign to fight terrorism, but was driven by oil companies and Jews in the U.S. and Israel.

2. Certain U.S. government officials planned the attacks of September 11, 2001 because they wanted the United States to go to war in the Middle East.

3. President Barack Obama was not really born in the United States and does not have an authentic Hawaiian birth certificate.

4. The current financial crisis was secretly orchestrated by a small group of Wall Street bankers to extend the power of the Federal Reserve and further their control of the world's economy.

5. Vapor trails left by aircraft are actually chemical agents deliberately sprayed in a clandestine program directed by government officials.
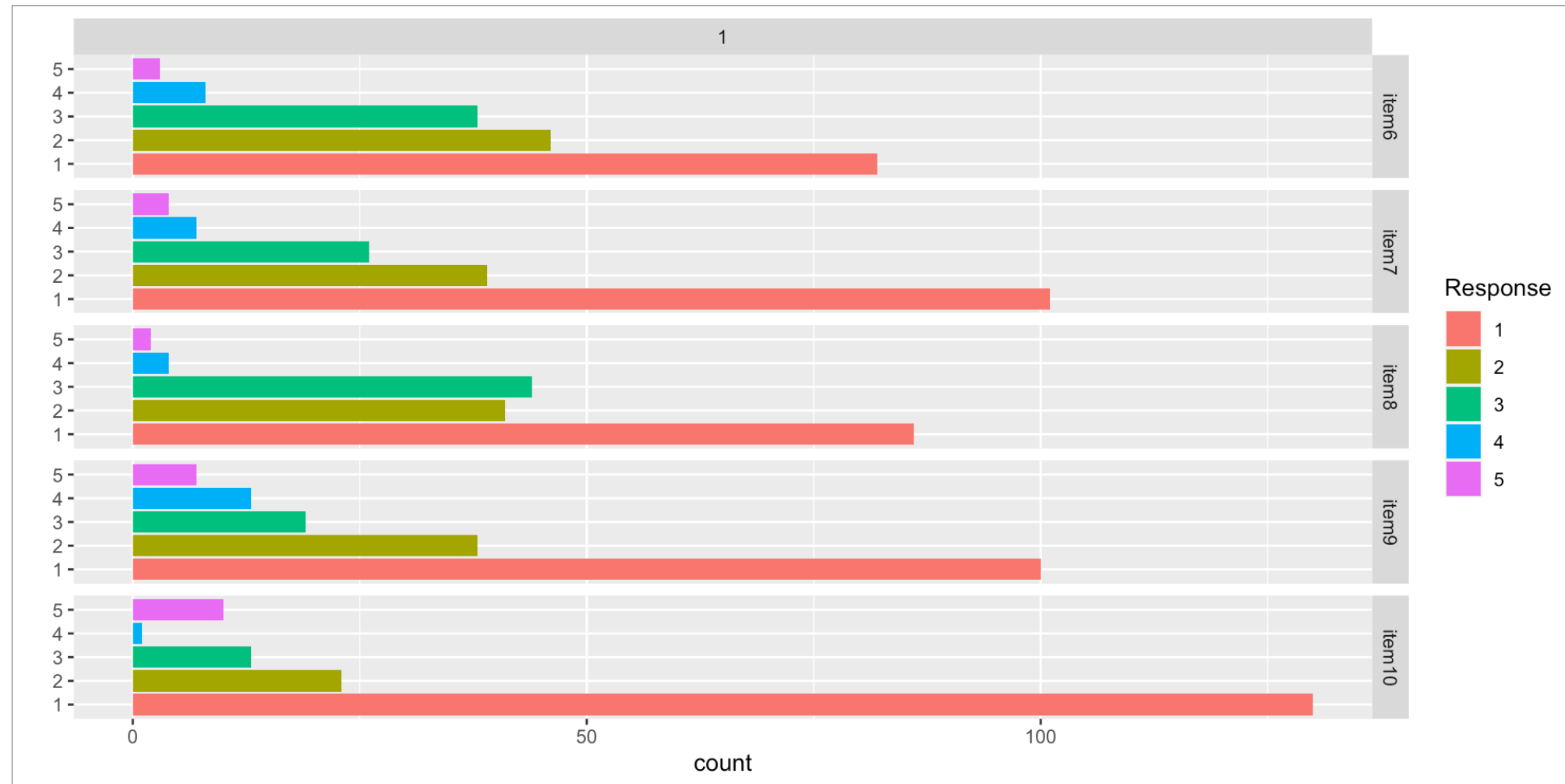
Conspiracy Theory Questions 6-10

Questions:

6. Billionaire George Soros is behind a hidden plot to destabilize the American government, take control of the media, and put the world under his control.

7. The U.S. government is mandating the switch to compact fluorescent light bulbs because such lights make people more obedient and easier to control.

8. Government officials are covertly Building a 12-lane "NAFTA superhighway" that runs from Mexico to Canada through America's heartland.

9. Government officials purposely developed and spread drugs like crack-cocaine and diseases like AIDS in order to destroy the African American community.

10. God sent Hurricane Katrina to punish America for its sins.

Data Visualization: Q1-Q5

Generalized Measurement Models: Modeling Observed Data

file:///Users/jonathantemplin/Library/CloudStorage/OneDrive-UniversityofIowa/02.%20OneDrive%20Teaching/PSQF%207375%20Bayesian%20P...

Data Visualization: Q6-Q10

Conspiracy Theories: Assumed Latent Variable

For today's lecture, we will assume each of the 10 items measures a single latent variable representing a **person's tendency to believe in conspiracy theories**

- We will denote this latent variable as θ_p for each person

    - p is the index for person (with $p = 1, \ldots, P$)

- We will assume this latent variable is:

    - Continuous

    - Normally distributed: $\theta_p \sim N(\mu_\theta, \sigma_\theta)$

        - We will make differing assumptions about the scale (the latent variable mean and standard deviation) to show how syntax works for either

- Across all people, we will denote the set of vector of latent variables as

$\boldsymbol{\Theta} = [\theta_1, \ldots, \theta_P]^T$

Generalized Measurement Models: Modeling Observed Data

file:///Users/jonathantemplin/Library/CloudStorage/OneDrive-UniversityofIowa/02.%20OneDrive%20Teaching/PSQF%207375%20Bayesian%20P...

Building Measurement Models

Measurement Model Analysis Steps

1. Specify model

2. Specify scale identification method for latent variables

3. Estimate model

4. Examine model-data fit

5. Iterate between steps 1-4 until adequate fit is achieved

Measurement Model Auxiliary Components

6. Score estimation (and secondary analyses with scores)

7. Item evaluation

8. Scale construction

9. Equating

10. Measurement invariance/differential item functioning

Modeling Observed Variables with Normal Distributions

Observed Variables with Normal Distributions

A psychometric model posits that one or more hypothesized latent variables predict a person's response to observed items

- Our hypothesized latent variable: Tendency to Believe in Conspiracies (θ_p)

    - One variable: Unidimensional

- Each observed variable (item response) is included in the model

- Today, we will assume each response follows a normal distribution

    - This is the assumption underlying confirmatory factor analysis (CFA) models

    - This assumption is tenuous at best

-

Normal Distribution: Linear Regression

As we saw in linear models, when an outcome variable (here Y_p) is assumed to follow a (conditional) normal distribution,
this places a linear regression-style model on the outcome:
For example, take the following linear regression:
$$Y_p = \beta_0 + \beta_1 X_p + e_p,$$
with $e_p \sim N(0, \sigma_e)$
This implies:
$$Y_p \sim N\left(\beta_0 + \beta_1 X_p, \sigma_e\right)$$
Where:

- The (conditional) mean of Y_p is $\beta_0 + \beta_1 X_p$

- The (residual) standard deviation of Y_p is σ_e

The Psychometric Model

For the psychometric model:

- We replace the observed variable X_p with the latent variable θ_p (for all observed variables)

$$Y_{p1} = \mu_1 + \lambda_1 \theta_p + e_{p,1}; \qquad e_{p,1} \sim N\left(0, \psi_1^2\right)$$

$$Y_{p2} = \mu_2 + \lambda_2 \theta_p + e_{p,2}; \qquad e_{p,2} \sim N\left(0, \psi_2^2\right)$$

$$Y_{p3} = \mu_3 + \lambda_3 \theta_p + e_{p,3}; \qquad e_{p,3} \sim N\left(0, \psi_3^2\right)$$

$$Y_{p4} = \mu_4 + \lambda_4 \theta_p + e_{p,4}; \qquad e_{p,4} \sim N\left(0, \psi_4^2\right)$$

$$Y_{p5} = \mu_5 + \lambda_5 \theta_p + e_{p,5}; \qquad e_{p,5} \sim N\left(0, \psi_5^2\right)$$

$$Y_{p6} = \mu_6 + \lambda_6 \theta_p + e_{p,6}; \qquad e_{p,6} \sim N\left(0, \psi_6^2\right)$$

$$Y_{p7} = \mu_7 + \lambda_7 \theta_p + e_{p,7}; \qquad e_{p,7} \sim N\left(0, \psi_7^2\right)$$

$$Y_{p8} = \mu_8 + \lambda_8 \theta_p + e_{p,8}; \qquad e_{p,8} \sim N\left(0, \psi_8^2\right)$$

$$Y_{p9} = \mu_9 + \lambda_9 \theta_p + e_{p,9}; \qquad e_{p,9} \sim N\left(0, \psi_9^2\right)$$

$$Y_{p10} = \mu_{10} + \lambda_{10} \theta_p + e_{p,10}; \quad e_{p,10} \sim N\left(0, \psi_{10}^2\right)$$

Measurement Model Parameters

For an item i the model is:
$$Y_{pi} = \mu_i + \lambda_i \theta_p + e_{p,i}; \quad e_{p,i} \sim N\left(0, \psi_i^2\right)$$
The parameters of the model use different notation from typical linear regression models and have different names (they are called item parameters)

- μ_i: Item intercept

  - The expected score on the item when $\theta_p = 0$

  - Similar to β_0

- λ_i: Factor loading or item discrimination

  - The change in the expected score of an item for a one-unit increase in θ_p

  - Similar to β_1

- ψ_i^2: Unique variance (Note: In Stan, we will have to specify ψ_e; the unique standard deviation)

  - The variance of the residuals (the expected score minus observed score)

  - Similar to residual variance σ_e^2

Model Specification

The set of equations on the previous slide formed step #1 of the Measurement Model Analysis Steps:

1. Specify Model

   The next step is:

2. Specify scale identification method for latent variables

   We will initially assume $\theta_p \sim N(0, 1)$, which allows us to estimate all item parameters of the model

- This is what we call a standardized latent variable

  - They are like Z-scores

Generalized Measurement Models: Modeling Observed Data

file:///Users/jonathantemplin/Library/CloudStorage/OneDrive-UniversityofIowa/02.%20OneDrive%20Teaching/PSQF%207375%20Bayesian%20P...

Implementing Normal Outcomes in Stan

Implementing Normal Outcomes in Stan

There are a few changes needed to make Stan estimate psychometric models with normal outcomes:

- The model (predictor) matrix cannot be used

  - This is because the latent variable will be sampled–so the model matrix cannot be formed as a constant

- The data will be imported as a matrix

  - More than one outcome means more than one column vector of data

- The parameters will be specified as vectors of each type

  - Each item will have its own set of parameters

  - Implications for the use of prior distributions

Generalized Measurement Models: Modeling Observed Data

file:///Users/jonathantemplin/Library/CloudStorage/OneDrive-UniversityofIowa/02.%20OneDrive%20Teaching/PSQF%207375%20Bayesian%20P...

Stan's `data {}` Block

```stan
 1  data {
 2    int<lower=0> nObs;                  // number of observations
 3    int<lower=0> nItems;                // number of items
 4    matrix[nObs, nItems] Y;             // item responses in a matrix
 5
 6    vector[nItems] meanMu;
 7    matrix[nItems, nItems] covMu;       // prior covariance matrix for coefficients
 8
 9    vector[nItems] meanLambda;          // prior mean vector for coefficients
10    matrix[nItems, nItems] covLambda;   // prior covariance matrix for coefficients
11
12    vector[nItems] psiRate;             // prior rate parameter for unique standard deviations
13  }
```

Stan's `parameters {}` Block

```
1  parameters {
2    vector[nObs] theta;              // the latent variables (one for each person)
3    vector[nItems] mu;               // the item intercepts (one for each item)
4    vector[nItems] lambda;           // the factor loadings/item discriminations (one for each item)
5    vector<lower=0>[nItems] psi;     // the unique standard deviations (one for each item)
6  }
```

Here, the parameterization of λ (factor loadings/discrimination parameters) can lead to problems in estimation

- The issue: $\lambda_i \theta_p = (-\lambda_i)(-\theta_p)$

  - Depending on the random starting values of each of these parameters (per chain), a given chain may converge to a different region

- To demonstrate (later), we will start with a different random number seed

  - Currently using 09102022: works fine

  - Change to 25102022: big problems

Generalized Measurement Models: Modeling Observed Data

file:///Users/jonathantemplin/Library/CloudStorage/OneDrive-UniversityofIowa/02.%20OneDrive%20Teaching/PSQF%207375%20Bayesian%20P...

Stan's `model` Block

```
 1  model {
 2
 3    lambda ~ multi_normal(meanLambda, covLambda); // Prior for item discrimination/factor loadings
 4    mu ~ multi_normal(meanMu, covMu);             // Prior for item intercepts
 5    psi ~ exponential(psiRate);                   // Prior for unique standard deviations
 6
 7    theta ~ normal(0, 1);                         // Prior for latent variable (with mean/sd specified)
 8
 9    for (item in 1:nItems){
10      Y[,item] ~ normal(mu[item] + lambda[item]*theta, psi[item]);
11    }
12
13  }
```

The loop here conducts the model, separately, for each item

- Assumption of conditional independence enables this

    - Non-independence would need multivariate normal model

- The item mean is set by the conditional mean of the model

    - The item SD is set by the unique variance parameter

- The loop puts each item's parameters into the equation

Choosing Prior Distributions for Parameters

There is not uniform agreement about the choices of prior distributions for item parameters

- We will use uninformative priors on each to begin

  - After first model analysis, we will discuss these choices and why they were made

- For now:

  - Item intercepts: $\mu_i \sim N\left(0, \sigma^2_{\mu_i} = 1000\right)$

  - Factor loadings/item discriminations: $\lambda_i \sim N(0, \sigma^2_{\lambda_i} = 1000)$

  - Unique standard deviations: $\psi_i \sim \text{exponential}\,(.01)$

Prior Density Function Plots

R's Data List Object

```r
1  # data dimensions
2  nObs = nrow(conspiracyItems)
3  nItems = ncol(conspiracyItems)
4
5  # item intercept hyperparameters
6  muMeanHyperParameter = 0
7  muMeanVecHP = rep(muMeanHyperParameter, nItems)
8
9  muVarianceHyperParameter = 1000
10 muCovarianceMatrixHP = diag(x = muVarianceMatrixHP, nrow = nItems)
11
12 # item discrimination/factor loading hyperparameters
13 lambdaMeanHyperParameter = 0
14 lambdaMeanVecHP = rep(lambdaMeanHyperParameter, nItems)
15
16 lambdaVarianceHyperParameter = 1000
17 lambdaCovarianceMatrixHP = diag(x = lambdaVarianceHyperParameter, nrow = nItems)
18
19 # unique standard deviation hyperparameters
20 psiRateHyperParameter = .01
21 psiRateVecHP = rep(.1, nItems)
22
23 modelCFA_data = list(
24   nObs = nObs,
25   nItems = nItems,
26   Y = conspiracyItems,
27   meanMu = muMeanVecHP,
28   covMu = muCovarianceMatrixHP,
29   meanLambda = lambdaMeanVecHP,
30   covLambda = lambdaCovarianceMatrixHP,
```

Running the Model In Stan

The Stan program takes longer to run than in linear models:

- Number of parameters = 207

    - 10 observed variables (with three item parameters each: μ_i, λ_i, and ψ_i)

    - 177 latent variables (one for each person: 177 parameters)

- Key syntax block is the observed variable definition:

```
1 Y[,item] ~ normal(mu[item] + lambda[item]*theta, psi[item]);
```

- cmdstanr samples call:

```
1 modelCFA_samples = modelCFA_stan$sample(
2   data = modelCFA_data,
3   seed = 09102022,
4   chains = 4,
5   parallel_chains = 4,
6   iter_warmup = 1000,
7   iter_sampling = 2000
8 )
```

- Note: Typically, longer chains are needed for larger models like this

    - These will become even longer when we use non-normal distributions for observed data

Model Results

- Checking convergence with $\hat{R}$ (PSRF):

```
[1] 1.006423
```

- Item Parameter Results:

```
# A tibble: 30 × 10
   variable    mean median     sd    mad    q5   q95  rhat ess_bulk ess_tail
   <chr>      <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
 1 mu[1]       2.37   2.37 0.0876 0.0862  2.23  2.51  1.00    1447.    2855.
 2 mu[2]       1.96   1.96 0.0850 0.0853  1.82  2.10  1.00     995.    2338.
 3 mu[3]       1.88   1.88 0.0825 0.0849  1.75  2.01  1.00    1081.    2746.
 4 mu[4]       2.01   2.01 0.0852 0.0856  1.87  2.15  1.00    1060.    2587.
 5 mu[5]       1.99   1.99 0.0842 0.0838  1.85  2.12  1.01     768.    1596.
 6 mu[6]       1.89   1.89 0.0775 0.0770  1.77  2.02  1.01     825.    1647.
 7 mu[7]       1.73   1.73 0.0766 0.0774  1.60  1.85  1.00    1025.    2523.
 8 mu[8]       1.84   1.84 0.0733 0.0739  1.72  1.97  1.01     803.    1771.
 9 mu[9]       1.81   1.81 0.0882 0.0883  1.67  1.96  1.00    1079.    2748.
10 mu[10]      1.52   1.52 0.0809 0.0802  1.39  1.65  1.00    1464.    3239.
11 lambda[1]   0.740  0.739 0.0838 0.0828 0.607 0.882 1.00    3429.    4389.
12 lambda[2]   0.870  0.867 0.0774 0.0778 0.748 1.00  1.00    2010.    4141.
13 lambda[3]   0.802  0.802 0.0766 0.0755 0.680 0.932 1.00    2249.    4009.
14 lambda[4]   0.843  0.840 0.0763 0.0760 0.723 0.973 1.00    2383.    4144.
15 lambda[5]   0.997  0.994 0.0715 0.0708 0.886 1.12  1.00    1439.    2565.
16 lambda[6]   0.899  0.896 0.0650 0.0651 0.796 1.01  1.00    1408.    2668.
17 lambda[7]   0.764  0.762 0.0704 0.0695 0.653 0.884 1.00    2473.    3577.
18 lambda[8]   0.854  0.851 0.0609 0.0603 0.757 0.959 1.00    1528.    2882.
19 lambda[9]   0.861  0.859 0.0806 0.0815 0.735 0.997 1.00    2196.    3286.
20 lambda[10]  0.673  0.669 0.0772 0.0759 0.550 0.804 1.00    3108.    4190.
21 psi[1]      0.893  0.891 0.0498 0.0492 0.815 0.978 1.00   15449.    6480.
22 psi[2]      0.734  0.732 0.0433 0.0427 0.667 0.809 1.00   12834.    6393.
23 psi[3]      0.782  0.780 0.0447 0.0442 0.712 0.859 1.00   14422.    6595.
```

Modeling Strategy vs. Didactic Strategy

At this point, one should investigate model fit of the model we just ran

- If the model does not fit, then all model parameters could be biased

  - Both item parameters and person parameters (θ_p)

- Moreover, the uncertainty accompanying each parameter (the posterior standard deviation) may also be biased

  - Especially bad for psychometric models as we quantify reliaiblity with these numbers

But, to teach generalized measurement models, we will first talk about differing models for observed data

- Different distributions

- Different parameterizations across the different distributions

Then we will discuss model fit methods

Investigating Item Parameters

One plot that can help provide information about the item parameters is the item characteristic curve (ICC)

- Not called this in CFA (but equivalent)

- The ICC is the plot of the expected value of the response conditional on the value of the latent traits, for a range of latent trait values

$$E\left(Y_{pi} \mid \theta_p\right) = \mu_i + \lambda_i \theta_p$$

- Because we have sampled values for each parameter, we can plot one ICC for each posterior draw

Posterior ICC Plots

Posterior Distribution for Item Parameters

Before moving onto the latent variables, let's note the posterior distribution of the item parameters (for a single item):

$f(\mu_i, \lambda_i, \psi_i \mid \boldsymbol{Y}) \propto f(\boldsymbol{Y} \mid \mu_i, \lambda_i, \psi_i) f(\mu_i, \lambda_i, \psi_i)$

$f(\mu_i, \lambda_i, \psi_i \mid \boldsymbol{Y})$ is the (joint) posterior distribution of the parameters for item i

- The distribution of the parameters conditional on the data

$f(\boldsymbol{Y} \mid \mu_i, \lambda_i, \psi_i)$ is the distribution we defined for our observed data:

$f(\boldsymbol{Y} \mid \mu_i, \lambda_i, \psi_i) \sim N\left(\mu_i + \lambda_i \theta_p, \psi_i\right)$

$f(\mu_i, \lambda_i, \psi_i)$ is the (joint) prior distribution for each of the parameters, which, are independent:

$f(\mu_i, \lambda_i, \psi_i) = f(\mu_i) f(\lambda_i) f(\psi_i)$

Investigating the Latent Variables

The estimated latent variables are then:

```
# A tibble: 177 × 10
    variable    mean  median    sd   mad      q5      q95  rhat ess_b…¹ ess_t…²
    <chr>      <dbl>   <dbl> <dbl> <dbl>   <dbl>    <dbl> <dbl>   <dbl>   <dbl>
  1 theta[1]   0.0182  0.0165 0.253 0.245 -0.399   0.441  1.00  6111.   4880.
  2 theta[2]   1.53    1.53   0.261 0.258  1.10    1.96   1.00  3708.   4902.
  3 theta[3]   1.70    1.70   0.266 0.265  1.27    2.14   1.00  3294.   4749.
  4 theta[4]  -0.931  -0.927  0.250 0.249 -1.35   -0.528  1.00  4841.   4979.
  5 theta[5]   0.0408  0.0369 0.244 0.239 -0.357   0.439  1.00  5235.   5090.
  6 theta[6]  -0.986  -0.985  0.254 0.252 -1.40   -0.569  1.00  4620.   4205.
  7 theta[7]  -0.342  -0.344  0.245 0.240 -0.746   0.0667 1.00  5411.   5294.
  8 theta[8]  -0.0569 -0.0542 0.250 0.252 -0.463   0.355  1.00  5289.   5432.
  9 theta[9]  -0.790  -0.789  0.250 0.249 -1.21   -0.385  1.00  5271.   4348.
 10 theta[10]  0.0373  0.0385 0.239 0.239 -0.352   0.432  1.00  5413.   5248.
 11 theta[11] -0.986  -0.983  0.250 0.247 -1.40   -0.582  1.00  4511.   4411.
 12 theta[12]  0.254   0.255  0.246 0.245 -0.150   0.661  1.00  5364.   5243.
 13 theta[13] -0.756  -0.756  0.249 0.249 -1.17   -0.347  1.00  4101.   4237.
 14 theta[14] -0.981  -0.975  0.255 0.254 -1.42   -0.575  1.00  4542.   4187.
 15 theta[15] -0.844  -0.844  0.251 0.251 -1.26   -0.435  1.00  5013.   5375.
 16 theta[16] -0.0694 -0.0680 0.245 0.247 -0.480   0.329  1.00  5559.   5647.
 17 theta[17] -0.0709 -0.0681 0.250 0.251 -0.487   0.343  1.00  5524.   5316.
 18 theta[18] -0.667  -0.665  0.247 0.251 -1.08   -0.270  1.00  5153.   4951.
 19 theta[19]  1.86    1.86   0.270 0.270  1.43    2.31   1.00  3579.   5417.
 20 theta[20] -0.932  -0.933  0.250 0.249 -1.34   -0.520  1.00  4524.   4578.
 21 theta[21] -0.257  -0.255  0.247 0.240 -0.665   0.153  1.00  5381.   4773.
 22 theta[22] -0.986  -0.983  0.252 0.254 -1.41   -0.571  1.00  4493.   5363.
 23 theta[23] -0.085  -0.082  0.249 0.249 -1.39   -0.576  1.00  4811.   5085.
```

Generalized Measurement Models: Modeling Observed Data

file:///Users/jonathantemplin/Library/CloudStorage/OneDrive-UniversityofIowa/02.%20OneDrive%20Teaching/PSQF%207375%20Bayesian%20P...

```
45 theta[45]  0.373   0.376  0.255 0.250 −0.0467  0.792   1.00  5716.  4510.
```

EAP Estimates of Latent Variables

    103 theta[10… −0.669 −0.668 0.250 0.246 −1.08   −0.261   1.00   4514.   4899.

Density of EAP Estimates

Density of All Posterior Draws

Comparing Two Posterior Distributions

Comparing EAP Estimate with Posterior SD

Comparing EAP Estimate with Sum Score

Posterior Distribution for Person Parameters

The posterior distribution of the person parameters (the latent variable; for a single person):
$$f(\theta_p \mid Y) \propto f\left(Y \mid \theta_p\right) f\left(\theta_p\right)$$
Here:

- $f(\theta_p \mid Y)$ is the posterior distribution of the latent variable conditional on the observed data

- $f(\_p)$ is the model (data) likelihood

$$f\left(Y \mid \theta_p\right) = \prod_{i=1}^{I} f(Y_i \mid \theta_p)$$

- $f(Y_i \mid \theta_p)$ is the observed data likelihood: $f(Y_i \mid \theta_p)\, N(\mu_i + \lambda_i \theta_p, \psi_i)$

- $f\left(\theta_p\right) \sim N(0, 1)$ is the prior likelihood for the latent variable θ_p

- Technical details: **https://jonathantemplin.com/wp-content/uploads/2022/10/sem15pre906_lecture11.pdf**

Measurement Model Estimation Fails

Recall: Stan's `parameters {}` Block

```
1  parameters {
2    vector[nObs] theta;              // the latent variables (one for each person)
3    vector[nItems] mu;              // the item intercepts (one for each item)
4    vector[nItems] lambda;         // the factor loadings/item discriminations (one for each item)
5    vector<lower=0>[nItems] psi;   // the unique standard deviations (one for each item)
6  }
```

Here, the parameterization of λ (factor loadings/discrimination parameters) can lead to problems in estimation

- The issue: $\lambda_i \theta_p = (-\lambda_i)(-\theta_p)$

  - Depending on the random starting values of each of these parameters (per chain), a given chain may converge to a different region

- To demonstrate (later), we will start with a different random number seed

  - Currently using 09102022: works fine

  - Change to 25102022: big problems

New Samples Syntax

Trying the same model with a different random number seed:

```
1  modelCFA_samplesFail = modelCFA_stan$sample(
2    data = modelCFA_data,
3    seed = 25102022,
4    chains = 4,
5    parallel_chains = 4,
6    iter_warmup = 2000,
7    iter_sampling = 2000
8  )
```

Convergence: FAIL

```
[1] 1.741764
```

Why Convergence Failed

The issue: $\lambda_i \theta_p = (-\lambda_i)(-\theta_p)$

```
# A tibble: 30 × 10
   variable      mean   median     sd    mad      q5    q95   rhat ess_b…¹ ess_t…²
   <chr>        <dbl>    <dbl>  <dbl>  <dbl>   <dbl>  <dbl>  <dbl>   <dbl>   <dbl>
 1 mu[1]       2.37e+0  2.37   0.0865 0.0852  2.22   2.51   1.00   1.32e3   2582.
 2 mu[2]       1.95e+0  1.95   0.0837 0.0838  1.82   2.09   1.00   9.35e2   2294.
 3 mu[3]       1.87e+0  1.87   0.0831 0.0812  1.74   2.01   1.00   1.14e3   2591.
 4 mu[4]       2.01e+0  2.01   0.0841 0.0842  1.87   2.15   1.00   9.70e2   2281.
 5 mu[5]       1.98e+0  1.98   0.0838 0.0840  1.84   2.12   1.00   7.40e2   1639.
 6 mu[6]       1.89e+0  1.89   0.0761 0.0775  1.77   2.02   1.00   7.12e2   1580.
 7 mu[7]       1.72e+0  1.72   0.0767 0.0754  1.60   1.85   1.00   9.66e2   2777.
 8 mu[8]       1.84e+0  1.84   0.0730 0.0720  1.72   1.96   1.00   7.00e2   1438.
 9 mu[9]       1.81e+0  1.81   0.0864 0.0865  1.66   1.95   1.00   1.00e3   2556.
10 mu[10]      1.52e+0  1.52   0.0812 0.0804  1.39   1.66   1.00   1.42e3   2979.
11 lambda[1]  -1.41e-3  0.00962 0.745 1.09   -0.849  0.845  1.74   6.08e0    117.
12 lambda[2]  -1.70e-3 -0.00212 0.874 1.29   -0.969  0.969  1.74   6.08e0    161.
13 lambda[3]  -1.89e-3 -0.0291  0.807 1.19   -0.906  0.906  1.74   6.06e0    104.
14 lambda[4]  -3.35e-4 -0.0119  0.848 1.25   -0.944  0.948  1.74   6.08e0    115.
15 lambda[5]  -1.54e-3  0.00194 1.00  1.48   -1.10   1.09   1.74   6.11e0    135.
16 lambda[6]  -1.22e-3 -0.00280 0.903 1.33   -0.986  0.984  1.74   6.09e0    119.
17 lambda[7]  -2.25e-3  0.0241  0.769 1.13   -0.858  0.855  1.74   6.07e0    105.
18 lambda[8]  -1.17e-3  0.00361 0.857 1.26   -0.938  0.935  1.74   6.11e0    132.
19 lambda[9]  -1.48e-3 -0.0127  0.867 1.28   -0.968  0.966  1.74   6.10e0    121.
20 lambda[10] -5.68e-4 -0.0454  0.678 0.995  -0.776  0.775  1.74   6.06e0    104.
21 psi[1]      8.92e-1  0.890   0.0511 0.0513  0.812  0.979  1.00   1.40e4   5818.
22 psi[2]      7.35e-1  0.732   0.0439 0.0437  0.666  0.811  1.00   1.24e4   6036.
23 psi[3]      7.81e-1  0.779   0.0451 0.0453  0.711  0.857  1.00   1.34e4   6747.
```

Posterior Trace Plots of λ

Posterior Density Plots of λ

Examining Latent Variables

```
# A tibble: 177 × 10
   variable       mean     median    sd    mad      q5    q95   rhat ess_bulk ess_t…¹
   <chr>         <dbl>      <dbl> <dbl> <dbl>   <dbl> <dbl>  <dbl>    <dbl>   <dbl>
 1 theta[1]   -0.00146   -0.00315 0.249 0.246 -0.408 0.403   1.00   4198.    4672.
 2 theta[2]    0.00245   -0.122   1.55  2.25  -1.85  1.86    1.73      6.07     98.4
 3 theta[3]    0.00266   -0.00440 1.72  2.52  -2.04  2.05    1.74      6.06    117.
 4 theta[4]    0.00206   -0.0296  0.959 1.37  -1.24  1.25    1.73      6.06    107.
 5 theta[5]   -0.00233   -0.00457 0.250 0.251 -0.409 0.410   1.02    201.     4934.
 6 theta[6]   -0.00499    0.0738  1.01  1.45  -1.30  1.29    1.73      6.08    124.
 7 theta[7]    0.00155   -0.00312 0.423 0.502 -0.655 0.664   1.61      6.59    106.
 8 theta[8]   -0.000405   0.000558 0.250 0.251 -0.409 0.407  1.03    108.     4176.
 9 theta[9]   -0.00238   -0.00591 0.826 1.16  -1.11  1.11    1.73      6.06    104.
10 theta[10]   0.000514  -0.00102 0.252 0.244 -0.411 0.412   1.02    238.     4244.
# … with 167 more rows, and abbreviated variable name ¹ess_tail
```

Posterior Trace Plots of θ

Posterior Density Plots of θ

Fixing Convergence

Stan allows starting values to be set via `cmdstanr`

- Documentation is very lacking, but with some trial and a lot of error, I will show you how

  Alternatively:

- Restrict λ to be positive

  - Truncates prior distribution with MVN

- Can also choose prior that has strictly positive range (like log-normal)

- Note: The restriction on the space of λ will not permit truely negative values

  - Not ideal as negative λ values are informative as a problem with data

```
1  parameters {
2    vector[nObs] theta;                // the latent variables (one for each person)
3    vector[nItems] mu;                 // the item intercepts (one for each item)
4    vector<lower=0>[nItems] lambda;         // the factor loadings/item discriminations (one for each item)
5    vector<lower=0>[nItems] psi;       // the unique standard deviations (one for each item)
6  }
```

Setting Starting Values in Stan

Starting values (initial values) are the first values used when an MCMC chain starts

- In Stan, by default, parameters are randomly started between -2 and 2

  - Bounded parameters are transformed so they are unbounded in the algorithm

- What we need:

  - Randomly start all λ parameters so that they converge to the $\lambda_i \theta_p$ mode

    - As opposed to the $(-\lambda_i)(-\theta_p)$ mode

cmdstanr Syntax for Initial Values

Add the init option to the $sample() function of the `cmdstanr` object:

```
 1  # set starting values for some of the parameters
 2  modelCFA_samples2fixed = modelCFA_stan$sample(
 3    data = modelCFA_data,
 4    seed = 25102022,
 5    chains = 4,
 6    parallel_chains = 4,
 7    iter_warmup = 2000,
 8    iter_sampling = 2000,
 9    init = function() list(lambda=rnorm(nItems, mean=10, sd=2))
10  )
```

The init option can be specified as a function, here, randomly starting each λ following a normal distribution

Initialization Process

See the lecture R syntax for information on how to confirm starting values are set

Final Results: Parameters

```
[1] 1.008906


# A tibble: 30 × 10
   variable    mean median     sd    mad    q5   q95  rhat ess_bulk ess_tail
   <chr>      <dbl>  <dbl>  <dbl>  <dbl> <dbl> <dbl> <dbl>    <dbl>    <dbl>
 1 mu[1]       2.37   2.37 0.0867 0.0877  2.23  2.51  1.00    1484.    3017.
 2 mu[2]       1.96   1.96 0.0855 0.0849  1.81  2.10  1.01    1023.    1964.
 3 mu[3]       1.88   1.88 0.0846 0.0828  1.74  2.02  1.00    1210.    2686.
 4 mu[4]       2.01   2.01 0.0865 0.0873  1.87  2.16  1.00    1168.    2681.
 5 mu[5]       1.99   1.99 0.0845 0.0842  1.85  2.12  1.01     787.    1846.
 6 mu[6]       1.90   1.90 0.0774 0.0751  1.77  2.02  1.01     818.    1814.
 7 mu[7]       1.72   1.72 0.0776 0.0763  1.60  1.85  1.01    1065.    2588.
 8 mu[8]       1.84   1.85 0.0731 0.0721  1.72  1.96  1.01     812.    1596.
 9 mu[9]       1.81   1.81 0.0876 0.0857  1.66  1.95  1.00    1148.    2308.
10 mu[10]      1.52   1.52 0.0817 0.0807  1.39  1.66  1.00    1507.    2944.
11 lambda[1]   0.740  0.738 0.0832 0.0822 0.608 0.880 1.00    3278.    4524.
12 lambda[2]   0.871  0.869 0.0757 0.0742 0.749 0.998 1.00    1916.    3797.
13 lambda[3]   0.802  0.800 0.0778 0.0778 0.678 0.931 1.00    2568.    4543.
14 lambda[4]   0.844  0.842 0.0766 0.0773 0.721 0.974 1.00    2120.    3592.
15 lambda[5]   0.998  0.996 0.0710 0.0707 0.885 1.12  1.00    1460.    2807.
16 lambda[6]   0.900  0.897 0.0640 0.0638 0.799 1.01  1.00    1449.    3056.
17 lambda[7]   0.765  0.762 0.0701 0.0699 0.654 0.883 1.00    2126.    3744.
18 lambda[8]   0.854  0.852 0.0611 0.0605 0.757 0.959 1.00    1392.    2754.
19 lambda[9]   0.862  0.860 0.0792 0.0780 0.737 0.996 1.00    2265.    4269.
20 lambda[10]  0.673  0.670 0.0773 0.0762 0.549 0.805 1.00    3194.    4112.
21 psi[1]      0.892  0.890 0.0498 0.0505 0.814 0.976 1.00   12915.    5995.
22 psi[2]      0.734  0.733 0.0424 0.0415 0.667 0.807 1.00   13155.    6267.
23 psi[3]      0.782  0.780 0.0455 0.0448 0.711 0.860 1.00   13300.    6363.
```

Comparing Results

Plot of all parameters across both algorithm runs:

Correlation of all parameters across both algorithm runs:

```
[1] 0.9999968
```

Wrapping Up

Wrapping Up

Today, we showed how to model observed data using a normal distribution

- Assumptions of Confirmatory Factor Analysis

  - Not appropriate for our data

  - May not be appropriate for many data sets

- We will have to keep our loading/discrimination parameters positive to ensure each chain converges to the same posterior mode

  - This will continue through the next types of data

- Next up, categorical distributions for observed data

  - More appropriate for these data as they are discrete categorical responses