Measurement Models and Identification

Introduction to Applied Bayesian Modeling

Christopher Hare

University of Georgia

ICPSR Summer Program 2014

Item Characteristic Curves and Utility

- The two-parameter IRT model:
  - $P_{ijy} = \Phi(\beta_j \theta_i - \alpha_j)$
- α_j = Difficulty parameter of item j
- β_j = Discrimination parameter of item j
- θ_i = Latent score for unit i

Item Characteristic Curves and Utility

- **Goal: measure students' math ability**

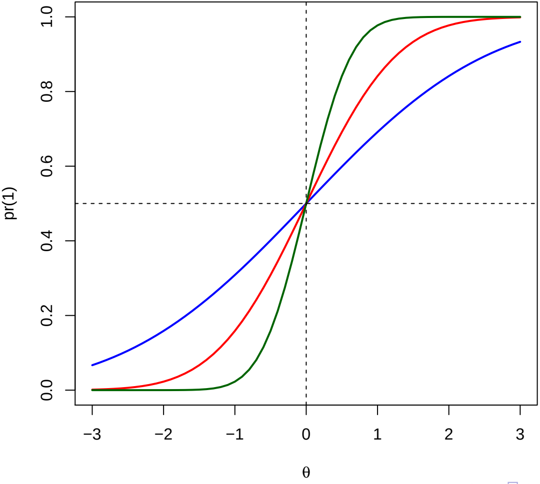
	Low β	High β
Low α	Spell "cat"	$3x = 6$
High α	Analysis of *Ulysses*	$\int_0^{\pi/2} \frac{\cos x}{e^{\sin x}} dx$

Item Characteristic Curves and Utility

- Plotting the ICC (or IRF) in R:

```
disc.mean <- c(0.5, 1, 2)
diff.mean <- c(0, 0, 0)
x <- seq(-3, 3, by = 0.1)
y1 <- pnorm((disc.mean[1] * x) - diff.mean[1])
y2 <- pnorm((disc.mean[2] * x) - diff.mean[2])
y3 <- pnorm((disc.mean[3] * x) - diff.mean[3])
plot(x, y1, ylim=c(0,1), type="l", lwd=2,
    col="blue", ylab="pr(1)", xlab=expression(theta))
lines(x, y2, lwd=2, col="red")
lines(x, y3, lwd=2, col="darkgreen")
abline(h=0.5, lty=2)
abline(v=0, lty=2)
```

Item Characteristic Curves and Utility

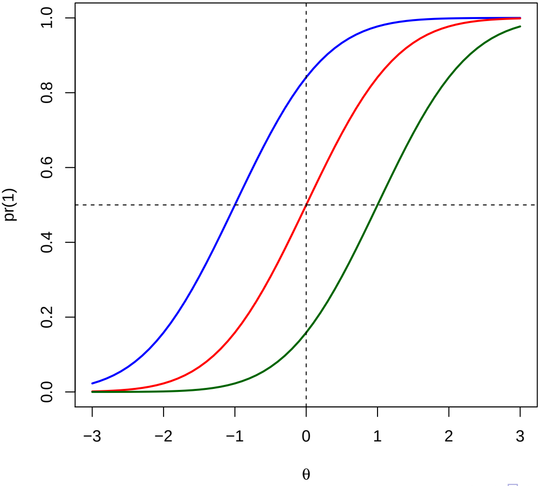


Item Characteristic Curves and Utility

- Can also shift the ICCs to the left and right (corresponding to more or less difficult items):

```
disc.mean <- c(1, 1, 1)
diff.mean <- c(-1, 0, 1)
x <- seq(-3, 3, by = 0.1)
y1 <- pnorm((disc.mean[1] * x) - diff.mean[1])
y2 <- pnorm((disc.mean[2] * x) - diff.mean[2])
y3 <- pnorm((disc.mean[3] * x) - diff.mean[3])
plot(x, y1, ylim=c(0,1), type="l", lwd=2,
    col="blue", ylab="pr(1)", xlab=expression(theta))
lines(x, y2, lwd=2, col="red")
lines(x, y3, lwd=2, col="darkgreen")
abline(h=0.5, lty=2)
abline(v=0, lty=2)
```

Item Characteristic Curves and Utility



Item Characteristic Curves and Utility

- The IRT model is so appealing to social scientists (well, at least some of us), because it is in line with a theory of individual choice behavior.
- Human behavior is rational, but probabilistic.
- When choices are very distant, individuals are less likely to make "errors."

Item Characteristic Curves and Utility

- Two-parameter IRT = Spatial voting with quadratic utility.

$$
\begin{aligned}
U_{ijy} &= -(\theta_i - O_{jy})^2 \\
U_{ijn} &= -(\theta_i - O_{jn})^2 \\
y^*_{ij} &= U_{ijy} - U_{ijn} + \varepsilon_{ij}
\end{aligned}
$$

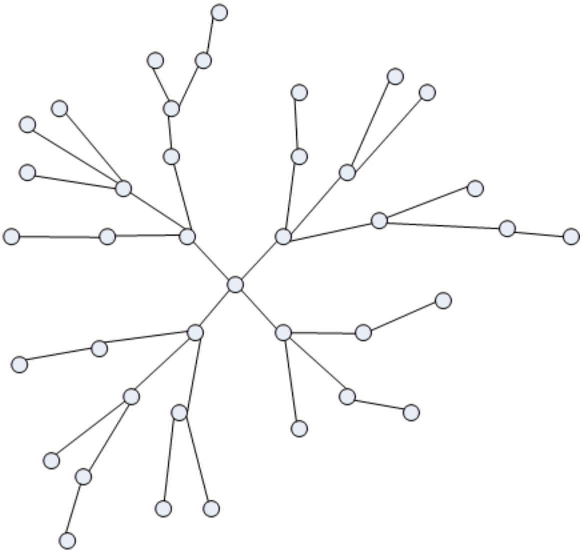
- By construction, let $\beta_j = 2(O_{jy} - O_{jn})$ and $\alpha_j = O^2_{jy} - O^2_{jn}$.
- Then:

$$
\begin{aligned}
y^*_{ij} &= \beta_{ij}\theta_i - \alpha_{ij} + \varepsilon_{ij} \\
\pi_{ijy} &= \pi(y^*_{ij} > 0) = \Phi(\beta_{ij}\theta_i - \alpha_{ij})
\end{aligned}
$$

Identification Issues in Measurement Models

- Why is identification a problem?
- In these types of models, what we are really estimating is the distance *between* the points (Jeffrey Lewis).

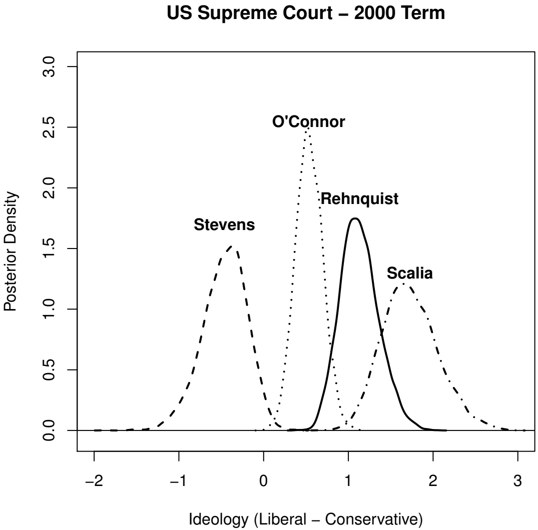
Identification Issues in Measurement Models



Identification Issues in Measurement Models

- Unique challenge for Bayesians: sampling from posterior distributions.

Identification Issues in Measurement Models



Identification Issues in Measurement Models

- The basic two-parameter IRT model (using the **Judges** example from Simon Jackman):

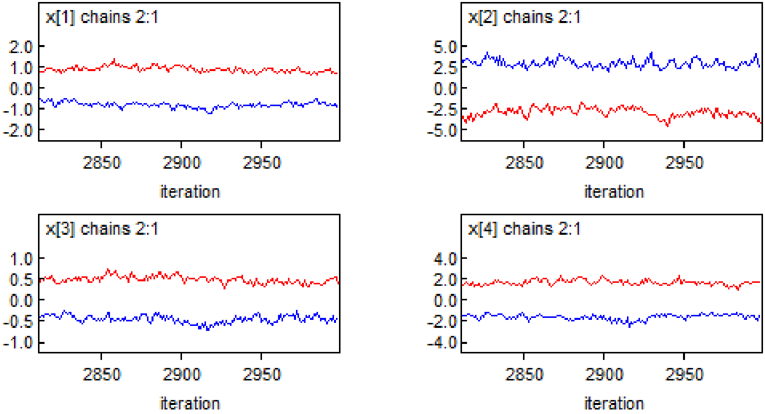
```
model{
for (i in 1:9){
  for (j in 1:228){
y[i,j] ~ dbern(p[i,j])
logit(p[i,j]) <- x[i]*beta[j,1] - beta[j,2]
     }
  }
```

Identification Issues in Measurement Models

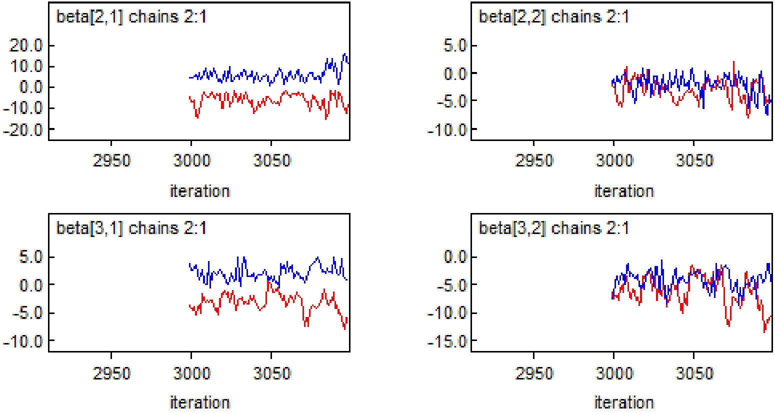
- Unidentified priors:

```
  ## priors
for (i in 1:9){
x[i] ~ dnorm(0.0, 1.0)
}

for(j in 1:228){
   beta[j,1:2] ~ dmnorm(b0[1:2],B0[1:2,1:2])
}
b0[1] <- 0 b0[2] <- 0
B0[1,1] <- .04 B0[2,2] <- .04
B0[1,2] <- 0 B0[2,1] <- 0

}
```

Identification Issues in Measurement Models



Identification Issues in Measurement Models



Identification Issues in Measurement Models

- How can we remedy this problem?
- The Kennedy-Helms constraint: constrain one θ to be -1, and another θ to be +1.
- Setting $s+1$ constraints (where s is the number of dimensions) is sufficient to achieve identification.

Identification Issues in Measurement Models

```
#priors
  x[1] ~ dnorm(0.0, 1.0)
  x[2] <- -1 ## Stevens
  x[3] ~ dnorm(0.0, 1.0)
  x[4] ~ dnorm(0.0, 1.0)
  x[5] ~ dnorm(0.0, 1.0)
  x[6] ~ dnorm(0.0, 1.0)
  x[7] <- 1 ## Thomas
  x[8] ~ dnorm(0.0, 1.0)
  x[9] ~ dnorm(0.0, 1.0)
```

Identification Issues in Measurement Models

- But, this is problematic, too.
- Bakker and Poole (2013): distances are no longer elastic, and uncertainty about the constrained legislator locations is transferred to others.

Identification Issues in Measurement Models

- Another solution: a series of sign constraints on the individual and/or item parameters to "freeze" the posteriors in certain quadrants.
- This *usually* achieves identification (especially in one dimension), but does not guarantee it.
- In two dimensions, Bakker and Poole (2013) propose setting one point at the origin to deal with shifts and pinning the second dimension coordinate of another point at 0 to deal with rotations.

Identification Issues in Measurement Models

```
#priors
  x[1] ~ dnorm(0.0, 1.0)
  x[2] ~ dnorm(0.0, 1.0)I(,0)
  x[3] ~ dnorm(0.0, 1.0)
  x[4] ~ dnorm(0.0, 1.0)
  x[5] ~ dnorm(0.0, 1.0)
  x[6] ~ dnorm(0.0, 1.0)
  x[7] ~ dnorm(0.0, 1.0)I(0,)
  x[8] ~ dnorm(0.0, 1.0)
  x[9] ~ dnorm(0.0, 1.0)
```

Identification Issues in Measurement Models

- Also, let's not neglect the item parameters.
- In the two-dimensional IRT model, we can achieve identification by setting one item to load *only* onto one dimension (Jackman 2001).

Identification Issues in Measurement Models

- Some practical advice: The Dark Art of Starts (Howard Rosenthal).
- Good starts are important–even critical–in these models. Why?

The Original Masters of Nasty Likelihood Surfaces

Identification Issues in Measurement Models

- This could include either hard-coding in or setting appropriate signs on (still dispersed) starting values:

```
> legislator.starts <- rnorm(N, 0, 1)
> legislator.starts[party=="R"] <-
        abs(legislator.starts[party=="R"])
> legislator.starts[party=="D"] <-
        -1 * abs(legislator.starts[party=="D"])
```

Identification Issues in Measurement Models

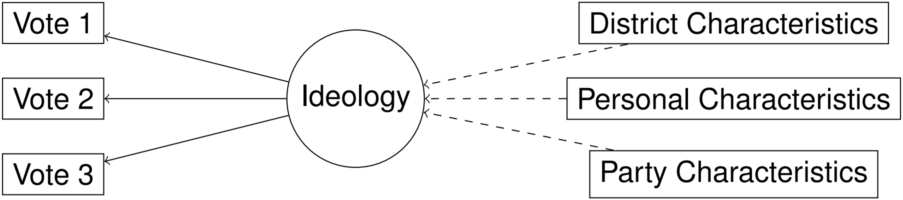
- Another popular option is to run some sort of maximum likelihood method and feed those results as the initial values.

Modifying the IRT Model

- The two-parameter IRT model: $P_{ijy} = \Phi(\beta_j \theta_i - \alpha_j)$

```
model{
for(i in 1:N){
for(j in 1:Q){
y[i,j] ~ dbern(pi[i,j])
probit(pi[i,j]) <- beta[j]*ideo[i] - alpha[j]
}}
## priors
for (i in 1:N){
ideo[i] ~ dnorm(0,.1)
}
for(j in 1:Q){
alpha[j] ~ dnorm(0,.1)
beta[j] ~ dnorm(0,.1)
}}
```

Modifying the IRT Model (MIMIC)



Modifying the IRT Model (MIMIC)

```
model{
for(i in 1:N){
for(j in 1:Q){
y[i,j] ~ dbern(pi[i,j])
probit(pi[i,j]) <- beta[j]*ideo[i] - alpha[j]
}}
## priors
for (i in 1:N){
ideo[i] ~ dnorm(0,.1)
}
for(j in 1:Q){
alpha[j] ~ dnorm(0,.1)
beta[j] ~ dnorm(0,.1)
}}
```

Modifying the IRT Model (MIMIC)

- The MIMIC (multiple indicators multiple causes) model:

```
## priors
for (i in 1:N){
ideo[i] ~ dnorm(mu[i],tau)
mu[i] <- a1 + b[1]*white[i] + b[2]*female[i]
    + b[3]*income[i] + b[4]*secular[i]
}
```

Modifying the IRT Model (Dynamic IRT)

- The use of "random-walk" priors to estimate dynamic measurement models introduced in political science by Martin and Quinn (2002).
- Random-walk priors are theoretically attractive because they are a compromise between two extremes: one that says parameters (like θ never change over time and the other that says parameters are entirely independent across time.
- The practical effect of random-walk priors is to smooth parameter estimates across time.

Modifying the IRT Model (Dynamic IRT)

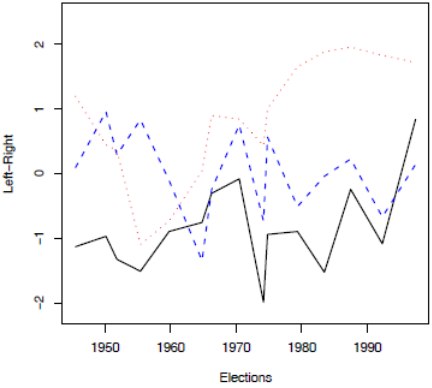
- The two-parameter IRT model:
  - $P_{ijy} = \Phi(\beta_j \theta_i - \alpha_j)$
- Modify to allow ideal points to be dynamic; index θ by time t ($t = 1, \ldots, T$).
- The two-parameter dynamic IRT model: $P_{ijyt} = \Phi(\beta_j \theta_{it} - \alpha_j)$.
- Random Walk Priors:

$$
\begin{aligned}
\theta_{i1} &\sim N(0, \tau_1) \\
\theta_{it} &\sim N(\theta_{i(t-1)}, \tau_2) \\
\tau_1 &\sim Gamma(1, 0.1) \\
\tau_2 &\sim Gamma(1, 0.1)
\end{aligned}
$$

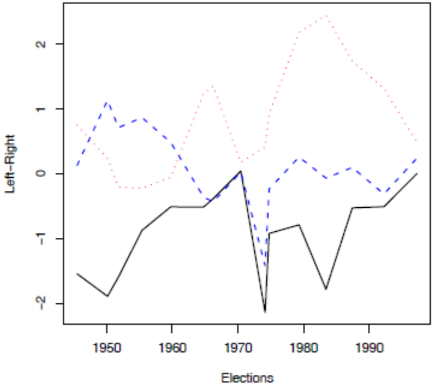
Modifying the IRT Model (Dynamic IRT)

```
model{
for(i in 1:N){
for(j in 1:Q){
for(t in 1:T){
y[i,j] ~ dbern(pi[i,j])
probit(pi[i,j]) <- beta[j]*ideo[i,t] - alpha[j]
}}
## priors
for (i in 1:N){
ideo[i,1] ~ dnorm(0, tau.Z[1])
for (t in 2:T){
ideo[i,t] ~ dnorm(ideo[i,t-1], tau.Z[2])
}}
for(j in 1:Q){
alpha[j] ~ dnorm(0, 0.1)
beta[j] ~ dnorm(0, 0.1)
}
tau.Z[1] ~ dgamma(1, 0.1)
tau.Z[2] ~ dgamma(1, 0.1)}
```

Modifying the IRT Model (Dynamic IRT)



(a) Manifesto



(b) Armstrong-Bakker

Modifying Measurement Models (BAM Scaling)

- Recall the factor model:
  - $y_{ij} = \beta_j \zeta_i + \varepsilon_{ij}$
- A useful way to think about this is that is specifies the DGP for the observed responses.
- Aldrich and McKelvey (1977) theorized a model in which respondents' issue scale placements are a function of the true positions of the stimuli, individual-specific distortions, and error.
- That is,
  - $z_{ij} = \alpha_i + \beta_i \zeta_j + \varepsilon_{ij}$

Modifying Measurement Models (BAM Scaling)

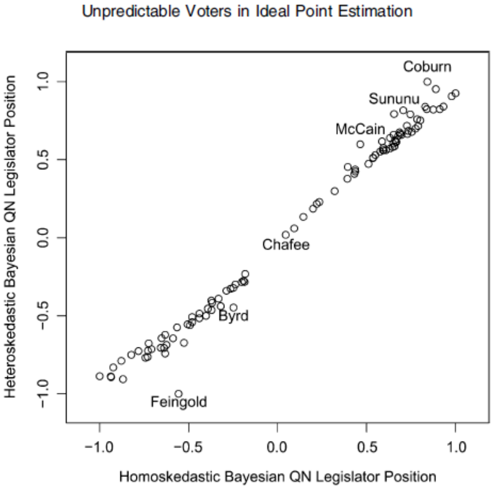
- Estimating the unknowns (α_i, β_i, and ζ_j) is straightforward in a Bayesian framework (and easy to code!):

```
model{
for(i in 1:N){  ## loop through respondents
        for(j in 1:q){  ## loop through stimuli
z[i,j] ~ dnorm(mu[i,j], tau[j])
mu[i,j] <- a[i] + b[i]*zeta[j]
}}
## priors on a and b
for(i in 1:N){
a[i] ~ dunif(-100,100)  ## priors on a and b
b[i] ~ dunif(-100,100)
}
for(j in 1:q){
tauj[j] ~ dgamma(.1,.1)  ## priors on tauj
zeta[j] ~ dnorm(0,1)  ## priors on zeta
}}
```

Modifying Measurement Models (Dealing with Heteroskedasticity)

- What about error variance in these models?
- The two-parameter IRT model:
  - $P_{ijy} = \Phi(\beta_j \theta_i - \alpha_j)$
- Lauderdale (2010) proposes adding a term that allows for unique error variance:
  - $P_{ijy} = \Phi(\frac{\beta_j \theta_i - \alpha_j}{\sigma_i})$
- Can also conceptualize as:
  - $P_{ijy} = \Phi(\beta_i \beta_j \theta_i - \alpha_j)$
- σ_i represents the degree to which the latent score influences the observed indicators (higher values $=$ more unpredictable).

Modifying Measurement Models (Dealing with Heteroskedasticity)

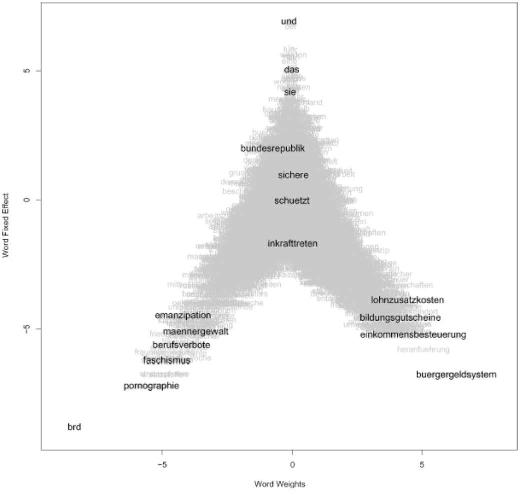


Modifying Measurement Models (Scaling Texts)

- Wordfish: Slapin and Proksch (2008).
- `http://www.wordfish.org/`
- Assume the number of times that party *i* mentions word *j* follows a Poisson distribution:
  - $y_{ij} = Poisson(\lambda_{ij})$
  - $\lambda_{ij} = \alpha_i + \Psi_j + \beta_j \theta_i$
- Where α_i and Ψ_j are party and word fixed effects, β_j measures the discrimination of word *j*, and θ_i is party *i*'s ideological position.
- Standard advantages of estimating this model in a Bayesian framework.

Modifying Measurement Models (Scaling Texts)

FIGURE 2 Word Weights vs. Word Fixed Effects. Left-Right Dimension, Germany 1990–2005 (Translations given in text)



Modifying Measurement Models (Scaling Texts)

TABLE 1 Top 10 Words Placing Parties on the Left and Right

	Top 10 Words Placing Parties on the...	
Dimension	**Left**	**Right**
Left-Right	Federal Republic of Germany (BRD)	general welfare payments (Bürgergeldsystem)
	immediate (sofortiger)	introduction (Heranführung)
	pornography (Pornographie)	income taxation (Einkommensbesteuerung)
	sexuality (Sexualität)	non-wage labor costs (Lohnzusatzkosten)
	substitute materials (Ersatzstoffen)	business location (Wirtschaftsstandort)
	stratosphere (Stratosphäre)	university of applied sciences (Fachhochschule)
	women's movement (Frauenbewegung)	education vouchers (Bildungsgutscheine)
	fascism (Faschismus)	mobility (Beweglichkeit)
	Two thirds world (Zweidrittelwelt)	peace tasks (Friedensaufgaben)
	established (etablierten)	protection (Protektion)