Computational methods for Bayesian semiparametric Item Response Theory models

Sally Paganin, joint work with Christopher J. Paciorek, Claudia Wehrhahn, Abel Rodriguez, Sophia Rabe-Hesketh, Perry de Valpine
sally.paganin@berkeley.edu @sampling_sally https://salleuska.github.io/
Berkeley UNIVERSITY OF CALIFORNIA

Background

Item Response Theory models

- data is typically individual answers to a set of questions/items
- $y_{ij} \in \{0,1\}$, item $i = 1,\dots,I$ from individual $j = 1,\dots,N$
- model the conditional probability $\pi_{ij} = \Pr(y_{ij} = 1|\lambda_i, \beta_i, \eta_j)$

2 parameters logistic model (2PL)

$$\text{logit}(\pi_{ij}) = \underbrace{\lambda_i(\eta_j - \beta_i)}_{\text{IRT parameterization}} = \underbrace{\lambda_i\eta_j + \gamma_i}_{\text{slope-intercept parameterization}}$$

- β_i **difficulty** parameter
- λ_i **discrimination** parameter → 1PL model if $\lambda_i = 1, \forall i = 1,\dots,I$
- η_j **individual latent trait** (ability, physical status)

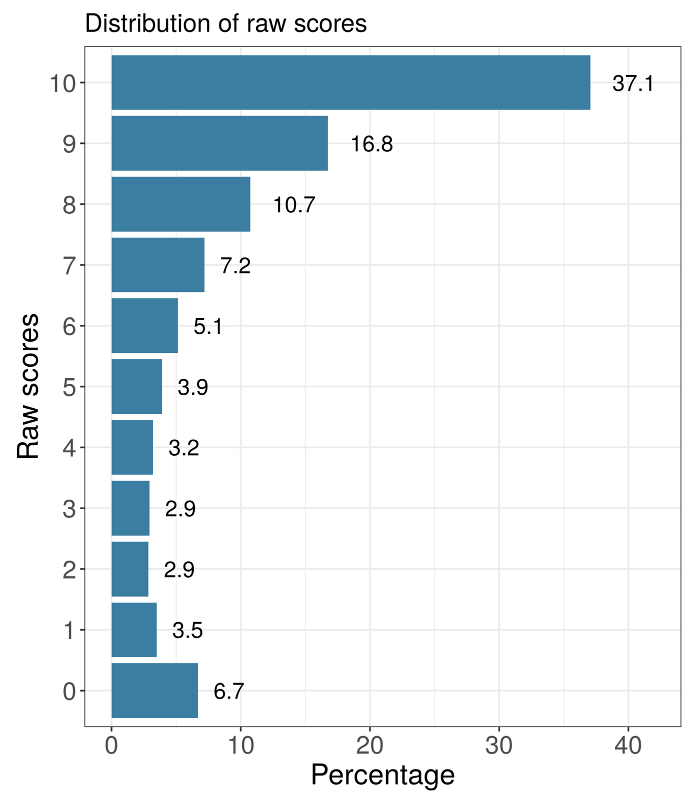
Traditional IRT models assume $\eta_j \sim \mathcal{N}(0,1)$ for $j = 1,\dots,N$

Is normality good for latent traits?

1996 England Health survey, N = 14,525, measuring **physical ability**

Q. Does your health now limit you in these activities? (yes, no)

1. Vigorous activities (e.g. sports)
2. Moderate activities (e.g. houseworks)
3. Lift/Carry (e.g. groceries)
4. Several stairs
5. One flight stairs
6. Bend/Kneel/Stoop
7. Walk more mile
8. Walk several blocks
9. Walk one block
10. Bathing/Dressing



Adding flexibility - BNP priors

We propose using a **Dirichlet Process mixture model** as a distribution for $\eta_j \sim G$

$$G = \int \mathcal{K}(\eta_j|\theta)F(d\theta), \quad F \sim \text{DP}(\alpha, G_0)$$

- $\mathcal{K}(\cdot|\theta)$ **probability kernel** - Normal distribution $\theta = \{\mu, \sigma^2\}$
- α **concentration parameter**
- G_0 **base distribution** for $\{\mu, \sigma^2\}$ → $G_0 \equiv \mathcal{N}(0, \sigma_0^2) \times \text{InvGamma}(\nu_1, \nu_2)$
- different representations & sampling algorithms
  → Stick-breaking process | Chinese restaurant process

Software for BNP IRT

NIMBLE

R-based software for hierarchical models

→ builds on and extends the BUGS language
→ provide functionalities for DP models (CRP, SB)
→ highly customizable (e.g. distributions, algorithms)
→ compile in C++ for fast execution

Easy model definition

```
code2PL <- nimbleCode({
  for(i in 1:I) {
    for(j in 1:N) {
      y[j, i] ~ dbern(pi[j, i])
      logit(pi[j, i]) <-  lambda[i]*(eta[j] - beta[i])
    }
  }
  for(i in 1:I) {
    log(lambda[i]) ~ dnorm(0.5, var = 0.5)
    beta[i] ~ dnorm(0,  var = 3)
  }
  for(j in 1:N) {
    eta[j] ~ dnorm(0, 1)
  }
})
```

2PL parametric model

$$y_{ij} \sim Bin(\pi_{ij})$$
$$\text{logit}(\pi_{ij}) = \lambda_i(\eta_j - \beta_i),$$
$$\log(\lambda_i) \sim \mathcal{N}(0.5, 0.5),$$
$$\beta_i \sim \mathcal{N}(0,3), \quad i = 1,\dots,I,$$
$$\eta_j \overset{iid}{\sim} \mathcal{N}(0,1), \quad j = 1,\dots,N.$$

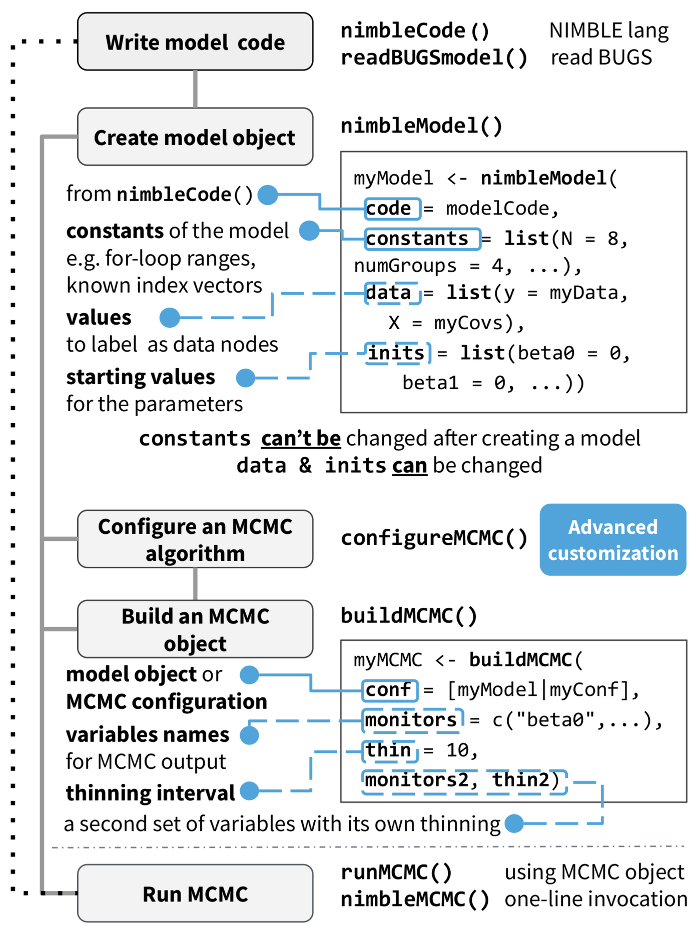
```
## CRP for clustering individual effects
zi[1:N] ~ dCRP(alpha, size = N)
alpha ~ dgamma(a, b)

## Mixture component drawn from the base measure
for(j in 1:N) {
  eta[j] ~ dnorm(mu[j], var = s2[j])
  mu[j] <- muTilde[zi[j]]
  s2[j] <- s2Tilde[zi[j]]
}

for(m in 1:N) {
  muTilde[m] ~ dnorm(0, var = s2_mu)
  s2Tilde[m] ~ dinvgamma(nu1, nu2)
}
```

2PL semiparametric model
→ extension to BNP priors with few lines of code

Workflow summary



NIMBLE

https://r-nimble.org/

```
# Create model
model <- nimbleModel(code2PL)
# Create and configure MCMC
mcmc <- configureMCMC(model)
mcmc$addSampler(...)

# Build MCMC
Rmcmc  <- buildMCMC(mcmc)
# Compile the model in C++
Cmodel <- compileNimble(model)
# Compile MCMC
Cmcmc  <- compileNimble(mcmc, project = "model")

# Get samples
samples <- runMCMC(Cmcmc, niter = 10000)
```

MCMC study for the 2PL model

Sampling strategy = model parametrization + identifiability constraints + sampling algorithm

Identifiability

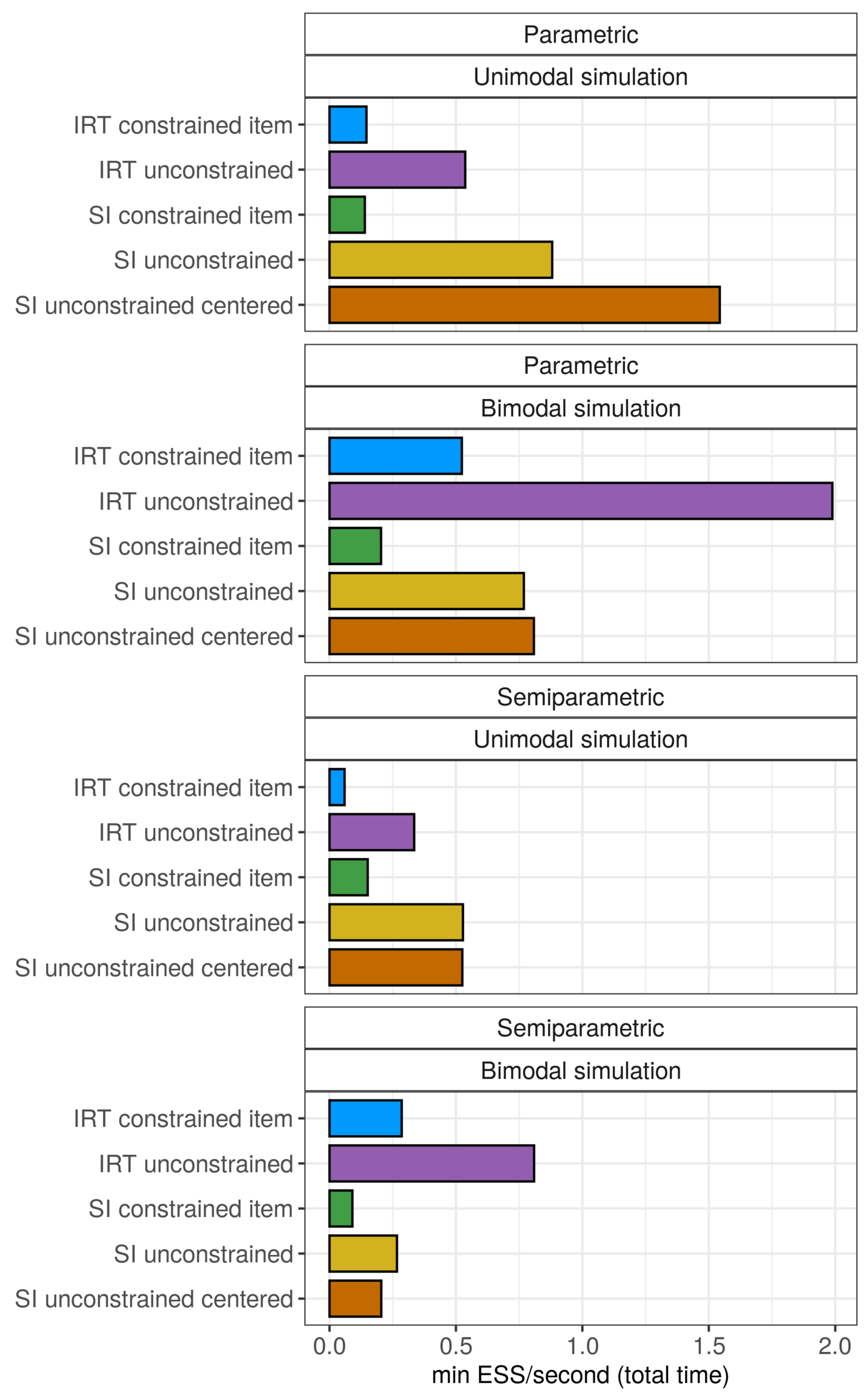
Identifiability can be addressed with different sets of constraints:

- on the abilities distribution $G \sim \mathcal{N}(0,1)$ → more complicated with DP
- on the item parameters, e.g. $\sum_{i=1}^{I} \beta_i = 0, \sum_{i=1}^{I} \log(\lambda_i) = 0$
  1. embedded in the model & and in the MCMC sampling
  2. post-processing MCMC samples, **parameter-expanded algorithms**

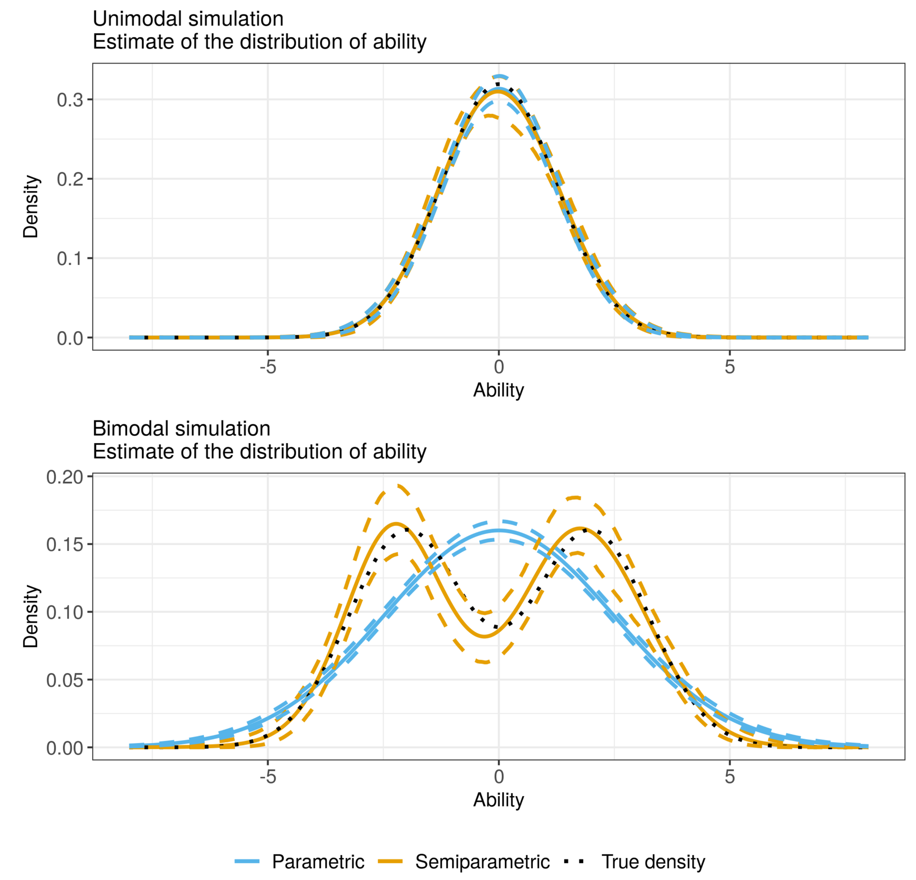
Sampling strategies

	Parametric		Semi-parametric	
Model constraints	slope-intercept	IRT	slope-intercept	IRT
1. Constrained item parameters	MH/conjugate	MH/conjugate*	MH/conjugate	MH/conjugate*
2. Unconstrained	MH/conjugate Centered	MH/conjugate	MH/conjugate Centered	MH/conjugate

Efficiency results



Inference results



- Efficiency changes with true distribution of the latent trait (IRT bimodal, SI unimodal)
- Misspecification of G can lead to biased estimates (also for item parameters)
- Computational cost for flexibility is reasonable in light of better inferential results