Nonparametric Deconvolution Models

Allison J.B. Chaney
Princeton University

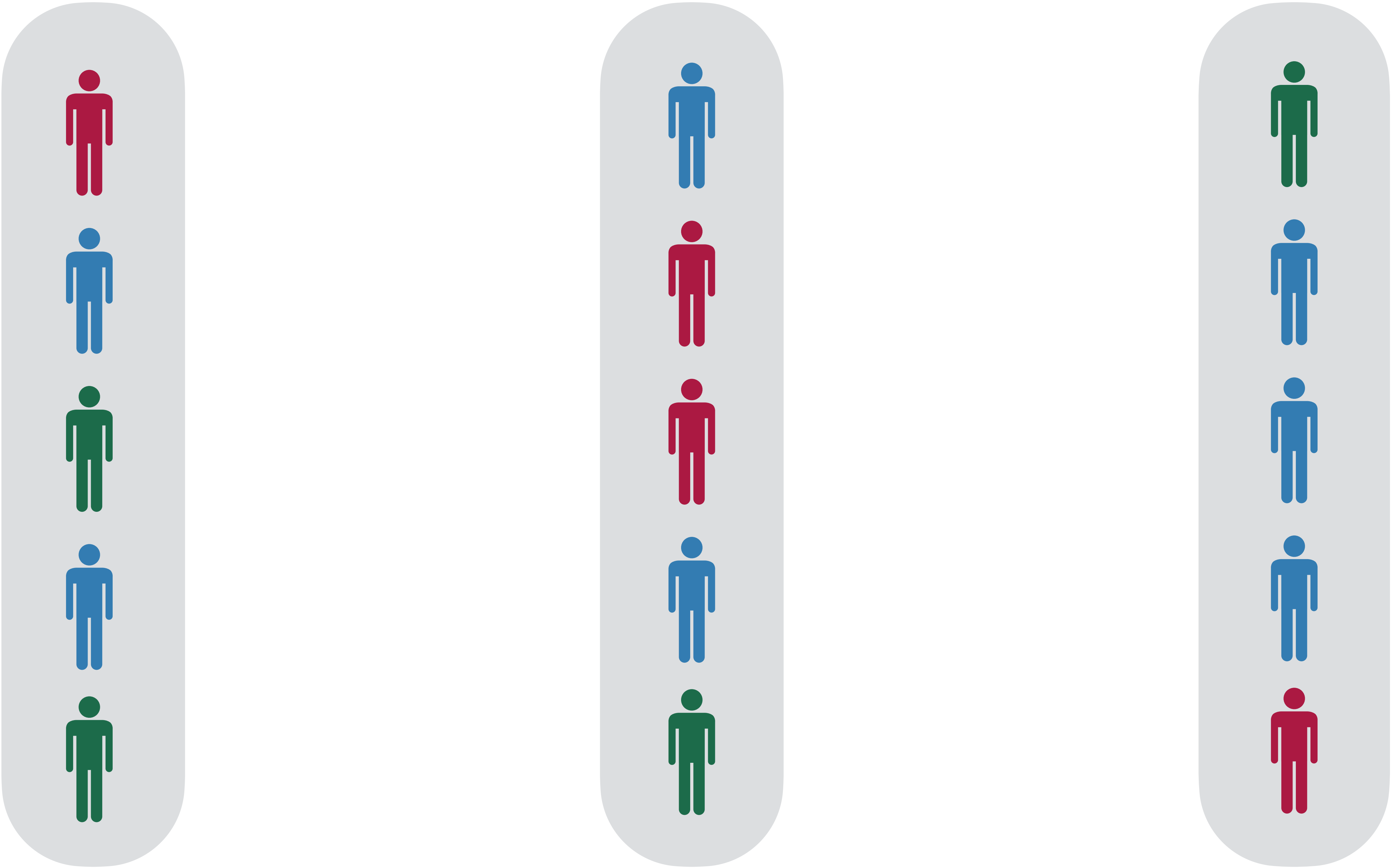
In collaboration with Barbara Engelhardt,
Archit Verma and Young-Suk Lee

Objective

Model collections of convolved data points

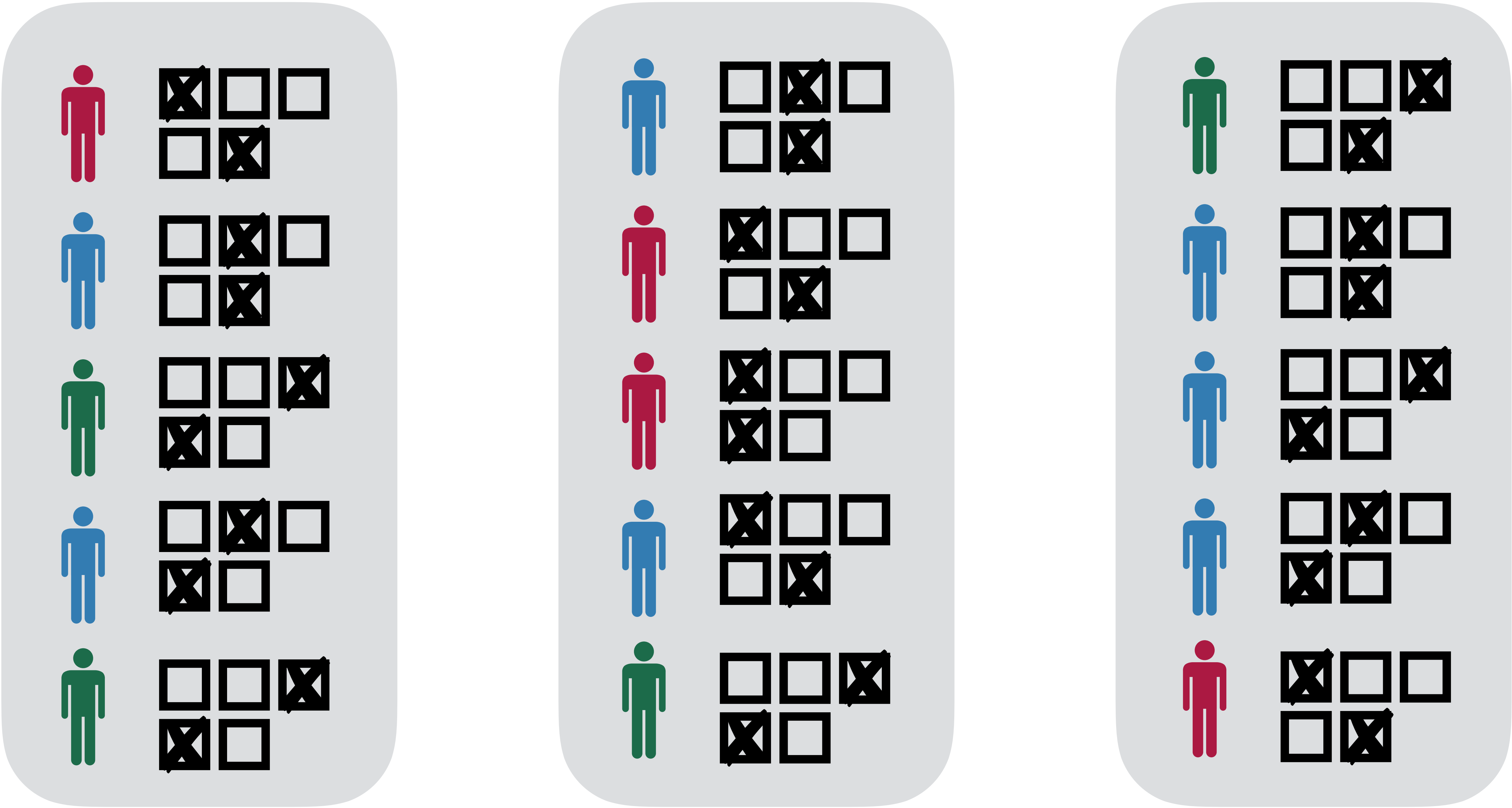
Objective

Model collections of convolved data points



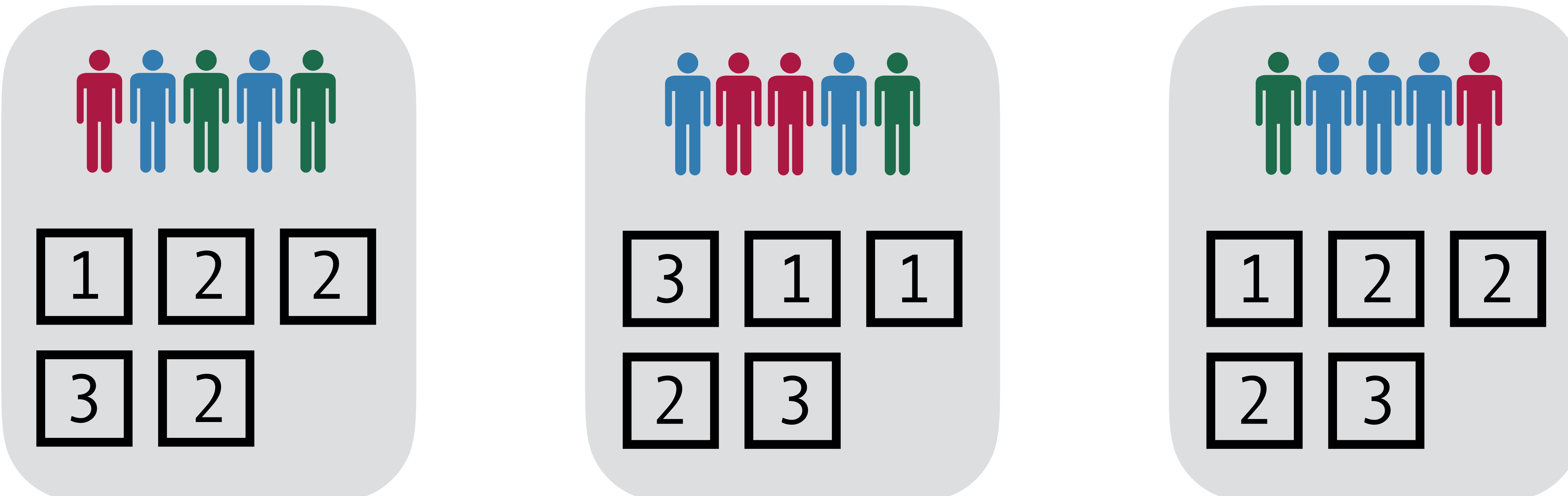
Objective

Model collections of convolved data points



Objective

Model collections of convolved data points



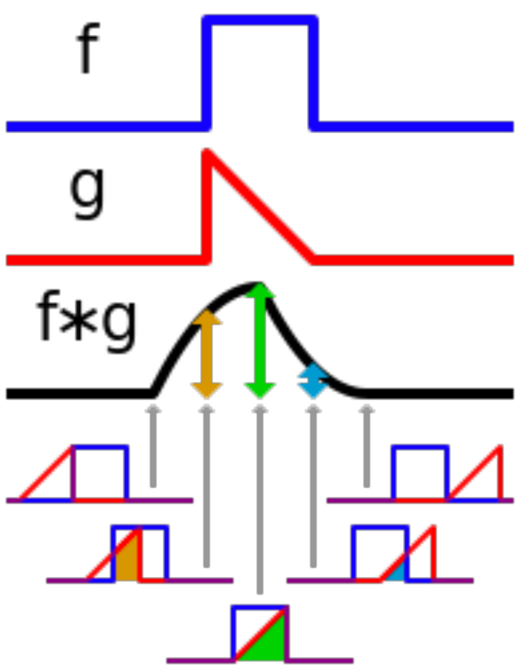
Objective

Model collections of convolved data points

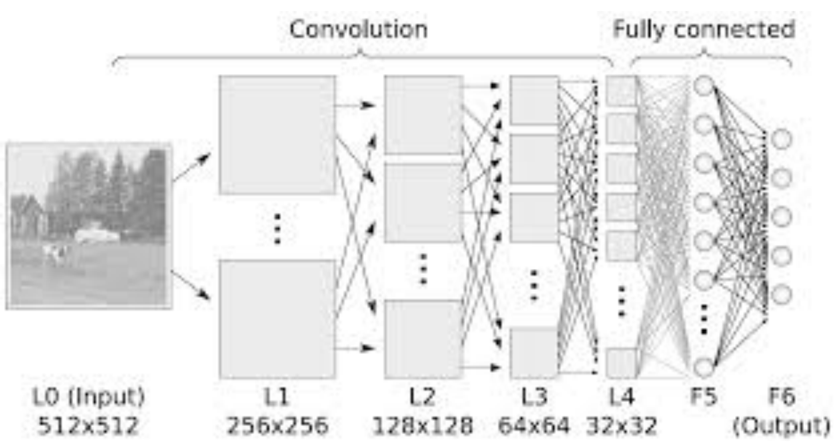
General	Voting	Bulk RNA-seq	Images
observation	district vote tally	sample	image
feature	issue or candidate	gene expression level	pixel
particle	individual voter	one cell	light particle
factor	voting cohort	cell type	visual pattern

“convolution”

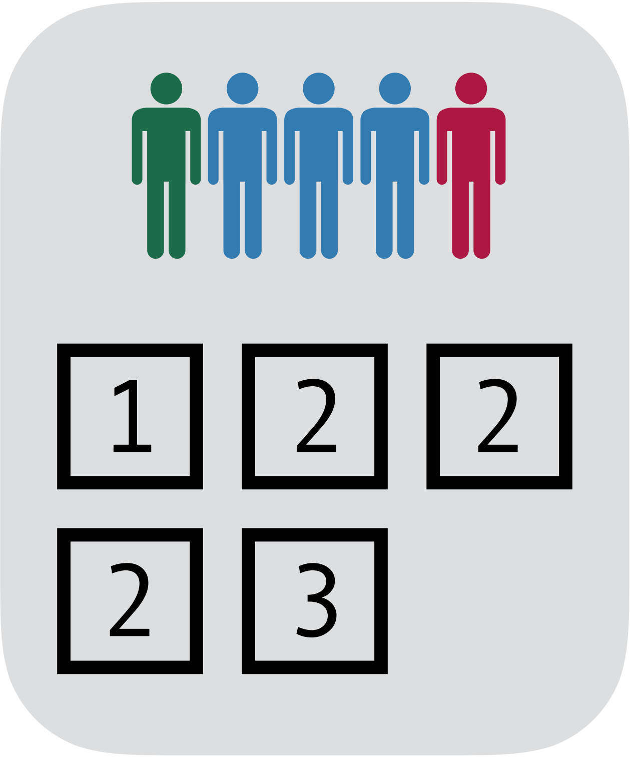
signal progressing



convolutional neural nets

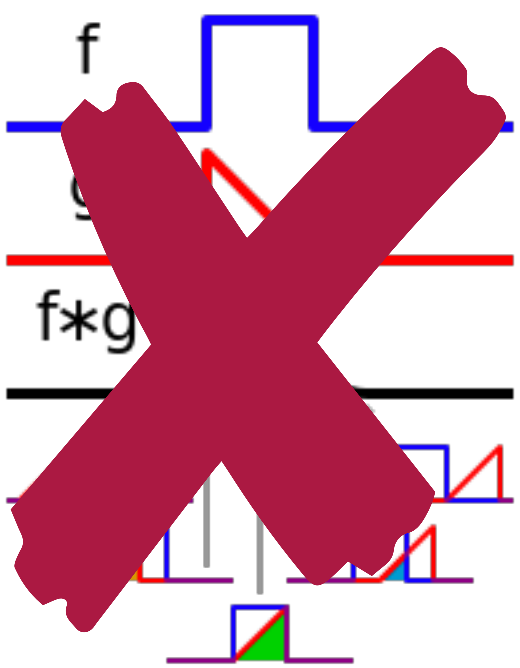


individual particles observed together

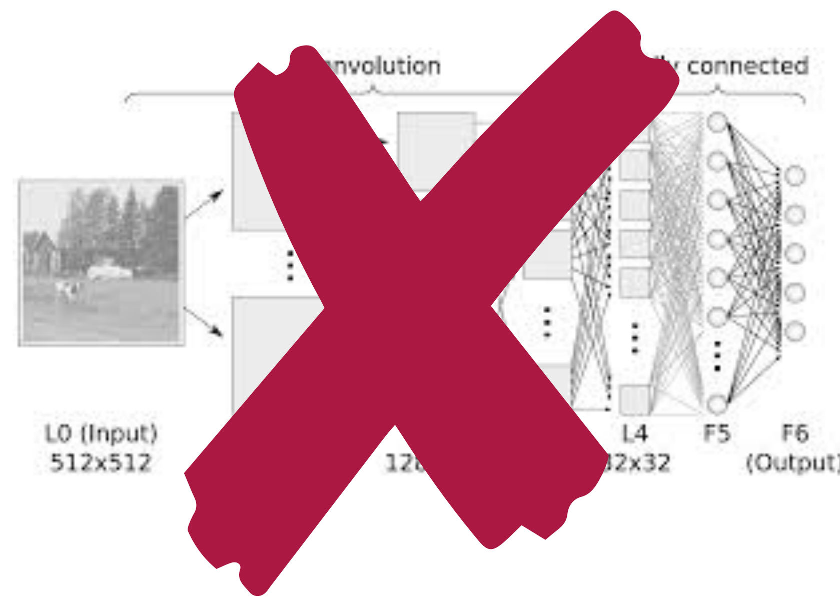


“convolution”

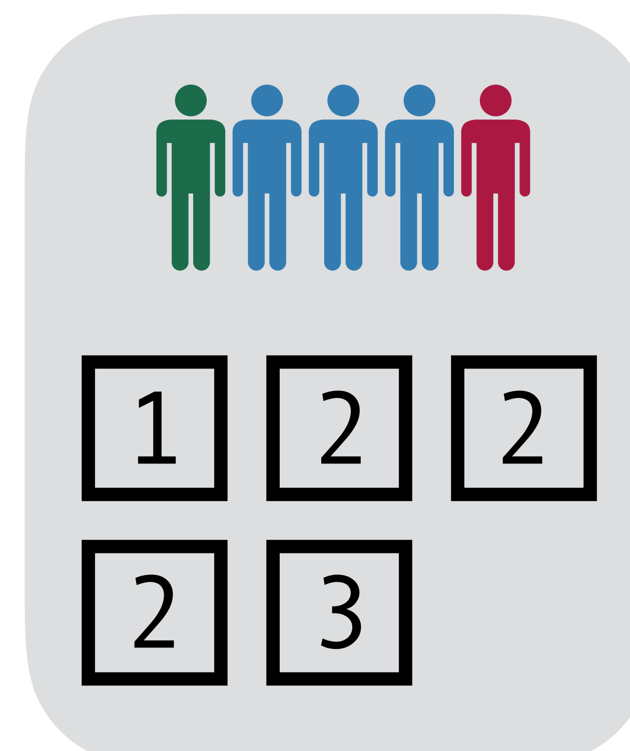
signal progressing



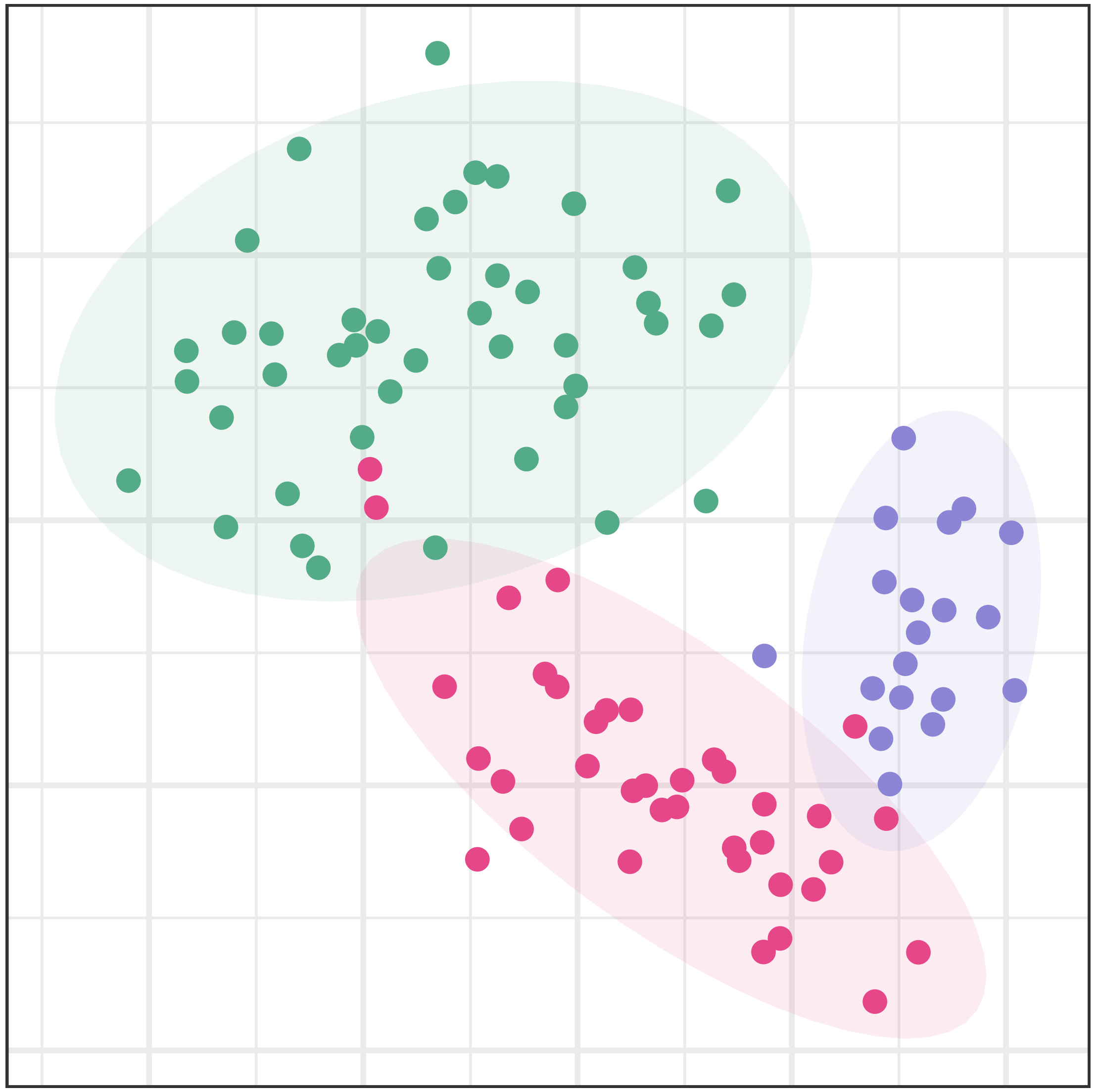
convolutional neural nets



individual particles observed together

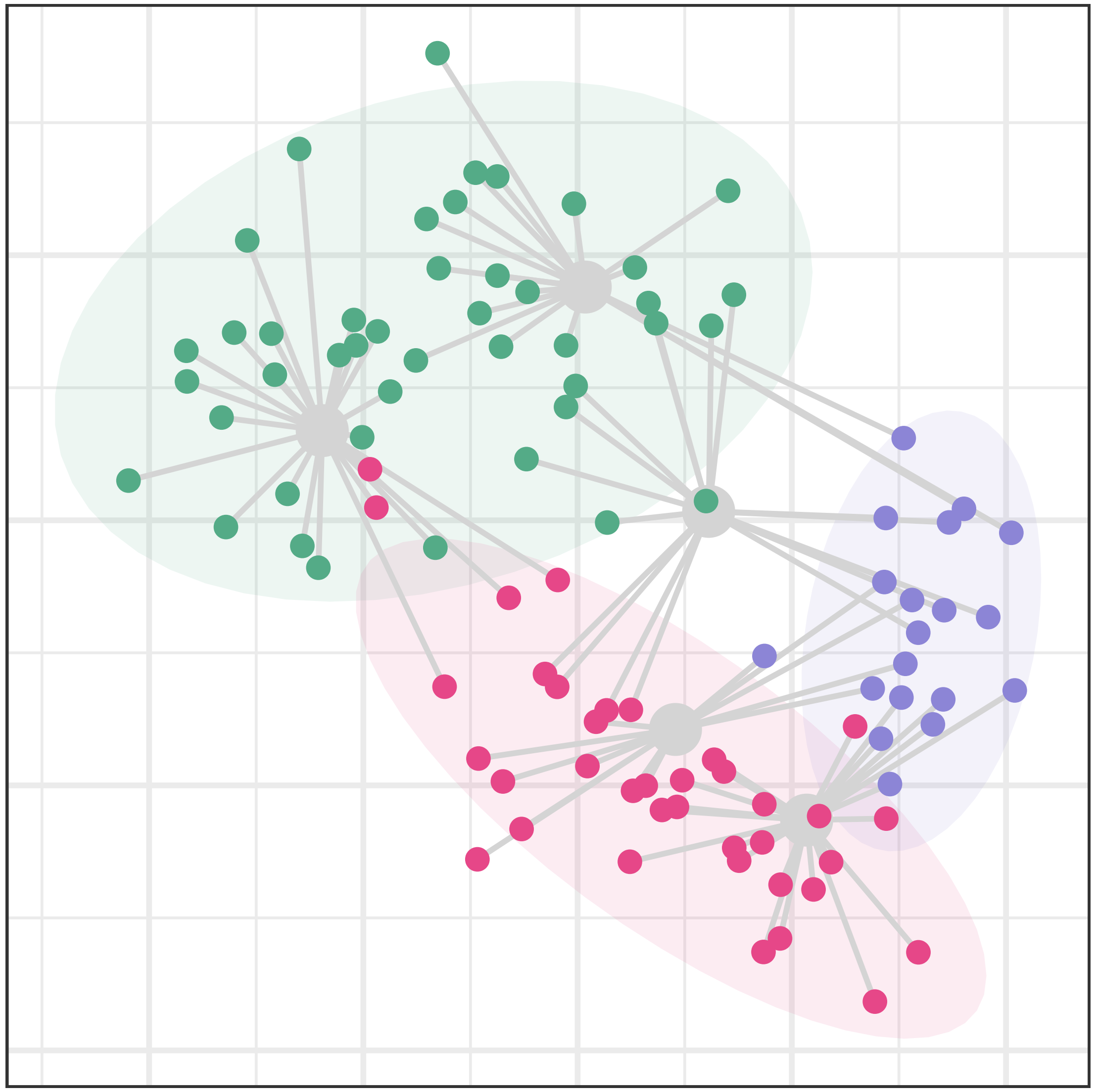


Related Models



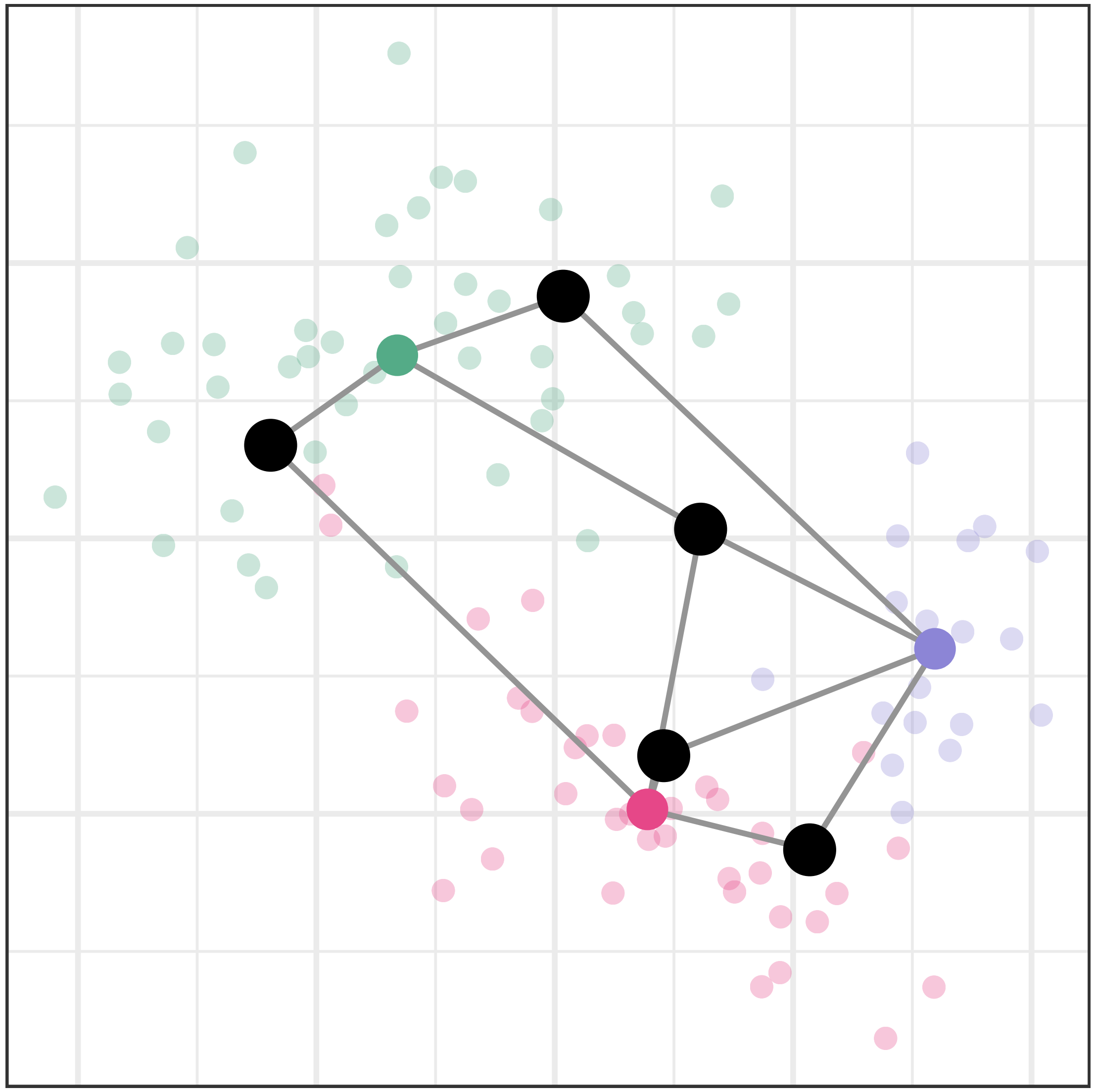
Mixture models assign each observation to one of K clusters, or factors.

Related Models



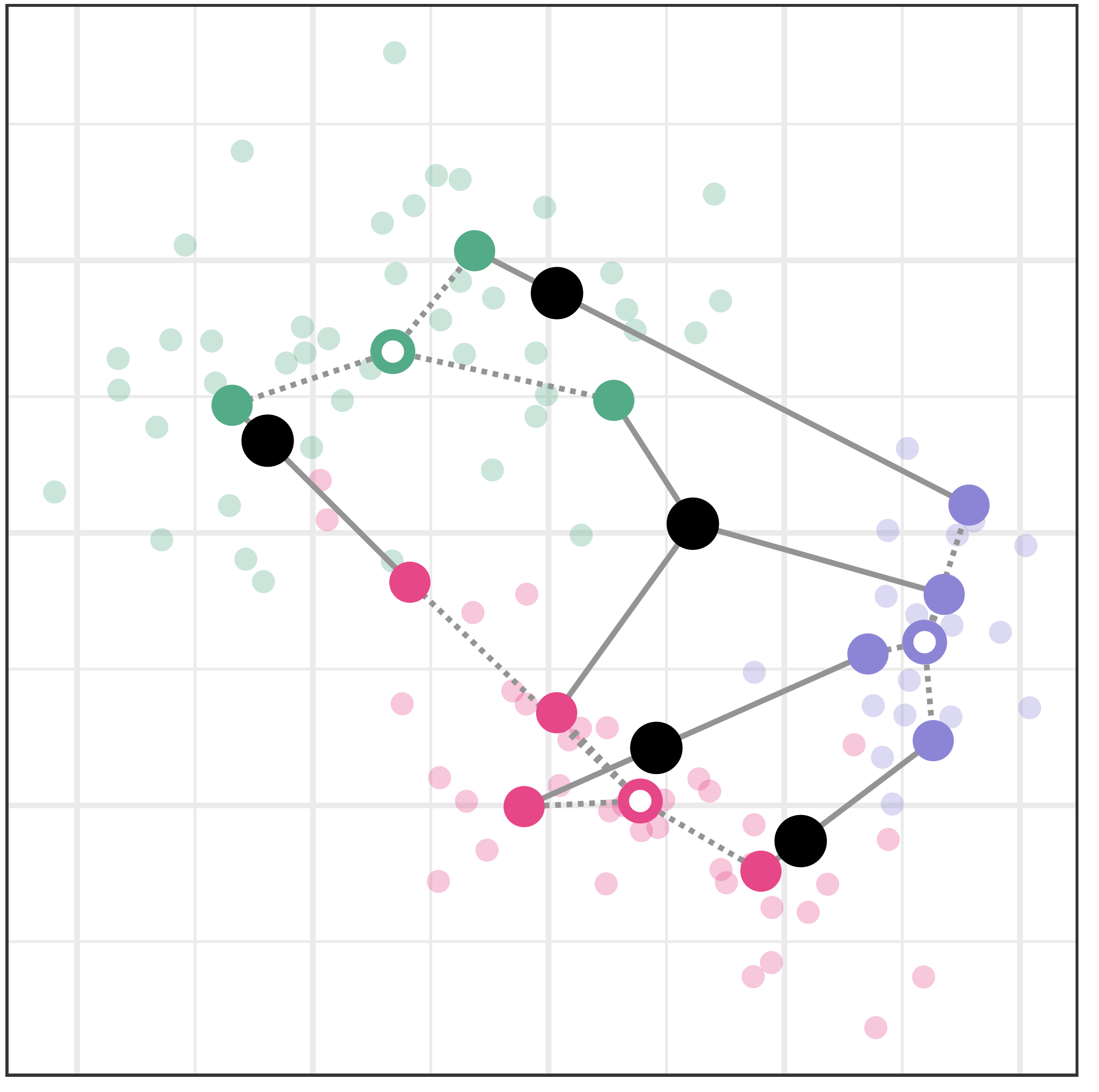
Admixture models represent groups of observations, each with its own mixture of K shared factors.

Related Models

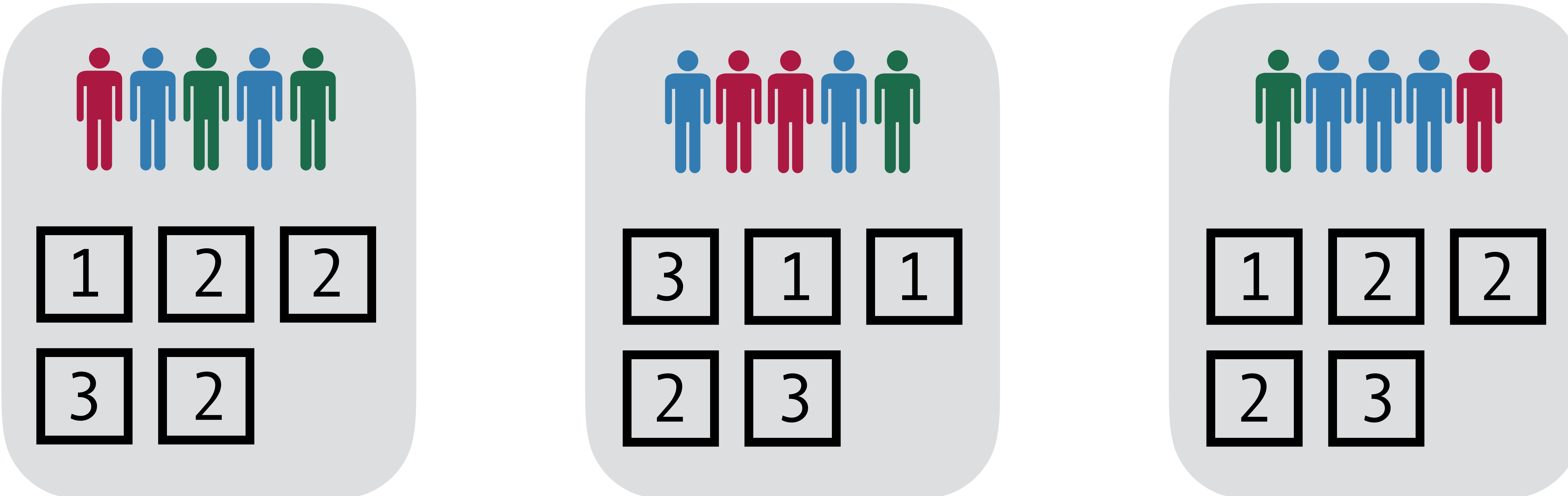


Decomposition models decompose observations into constituent parts by representing observations as a product between group representations and factor features.

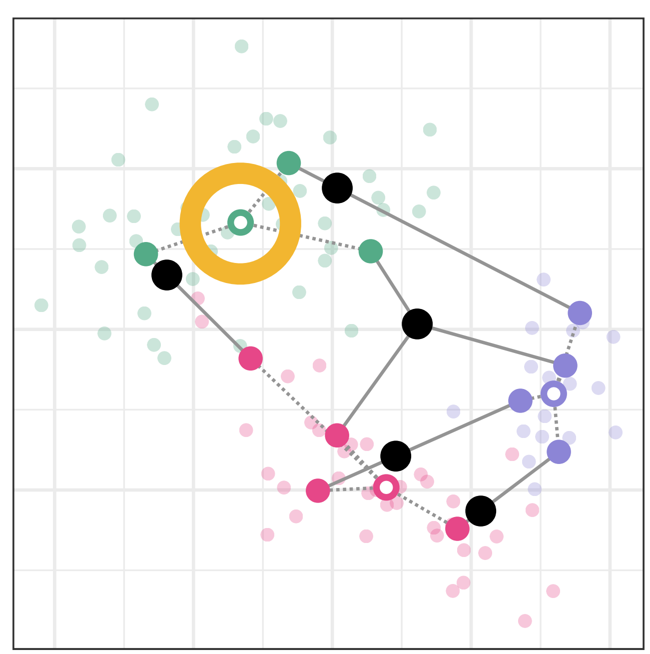
Our Model

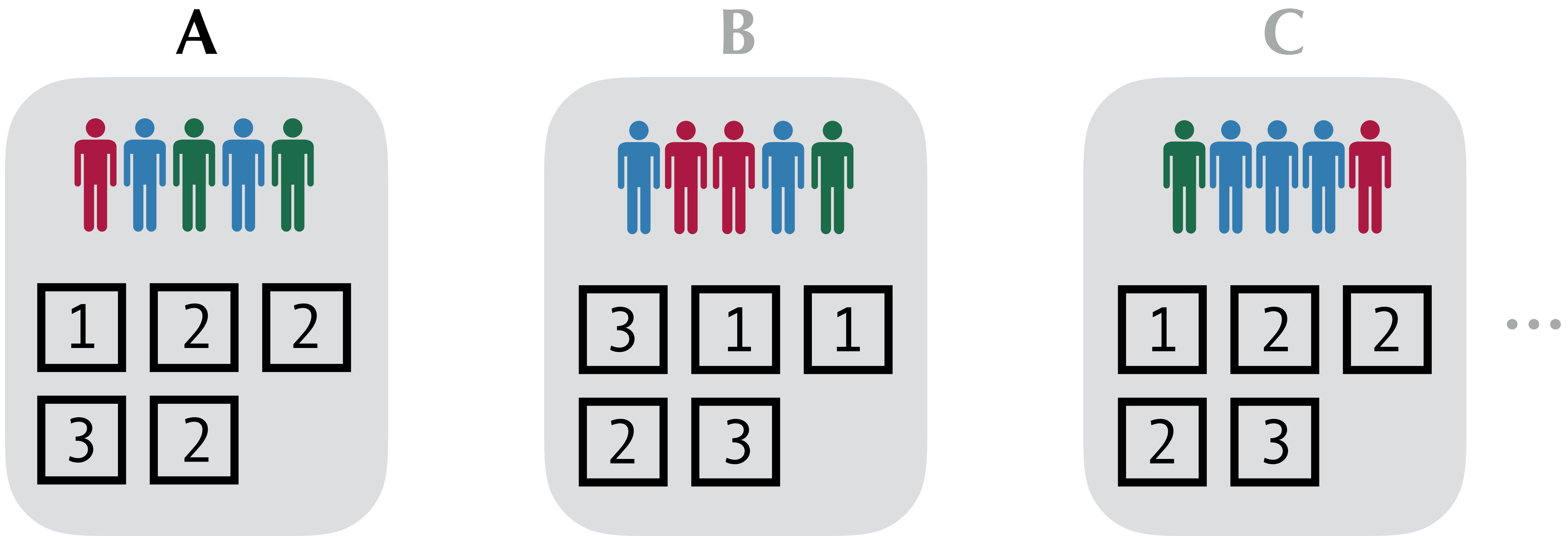


Deconvolution models (this work) similarly decompose, or deconvolve, observations into constituent parts, but also capture group-specific (or local) fluctuations in factor features.

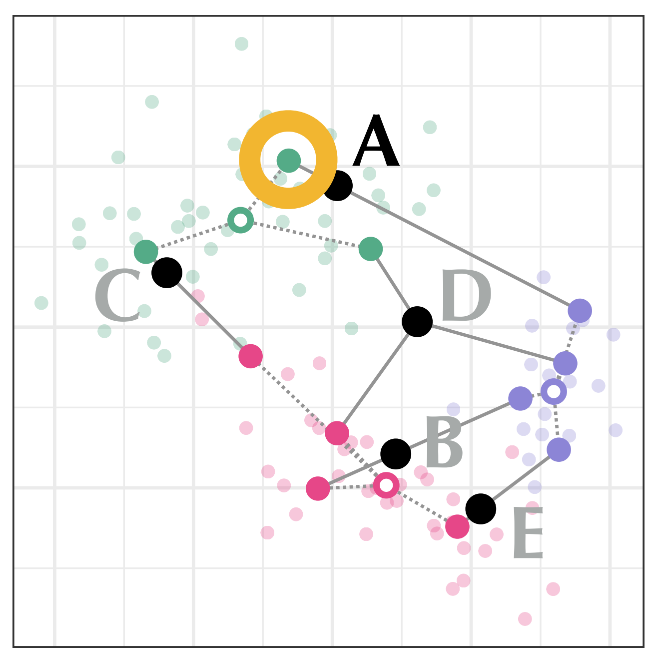


How do  usually vote?

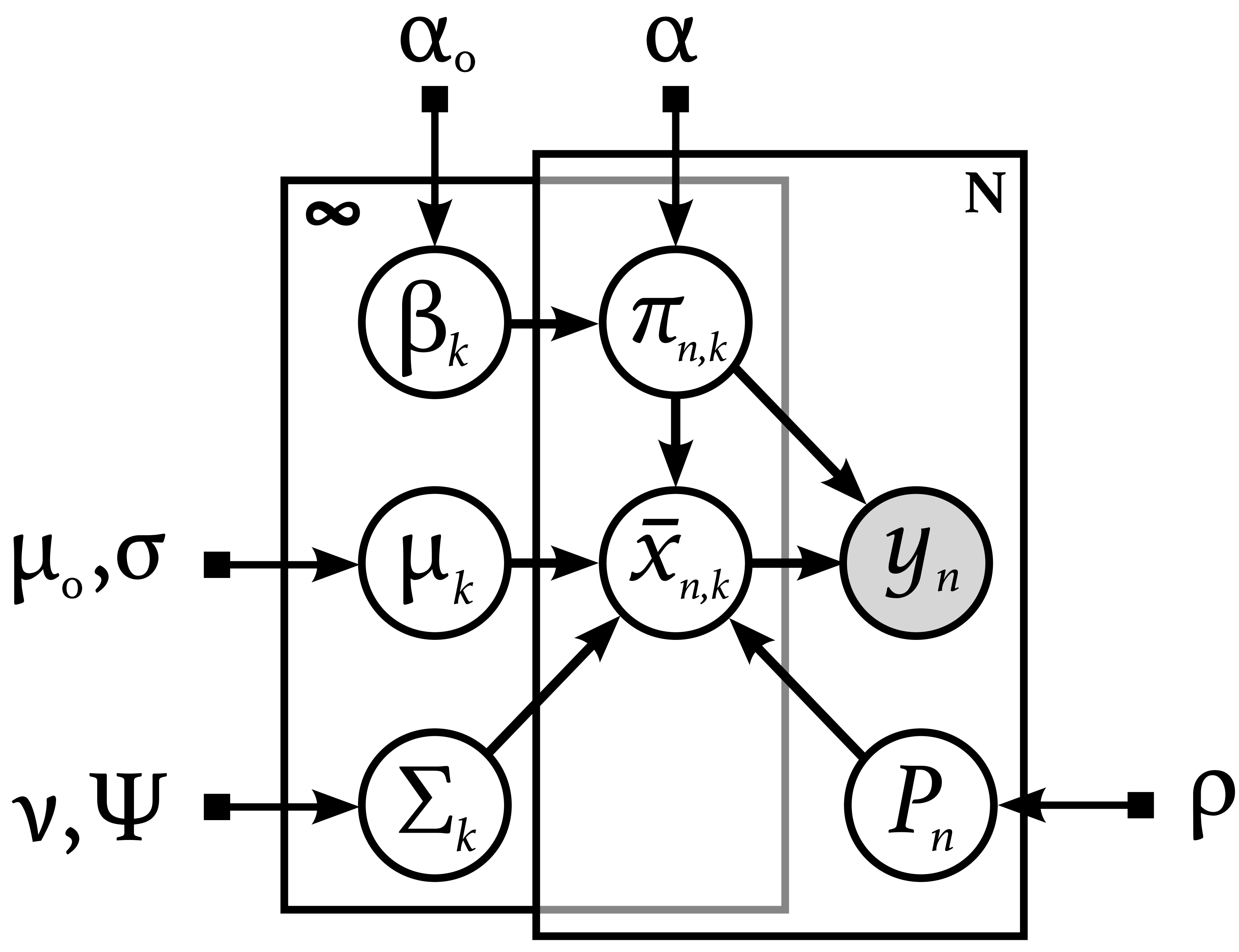




How do vote in district A?

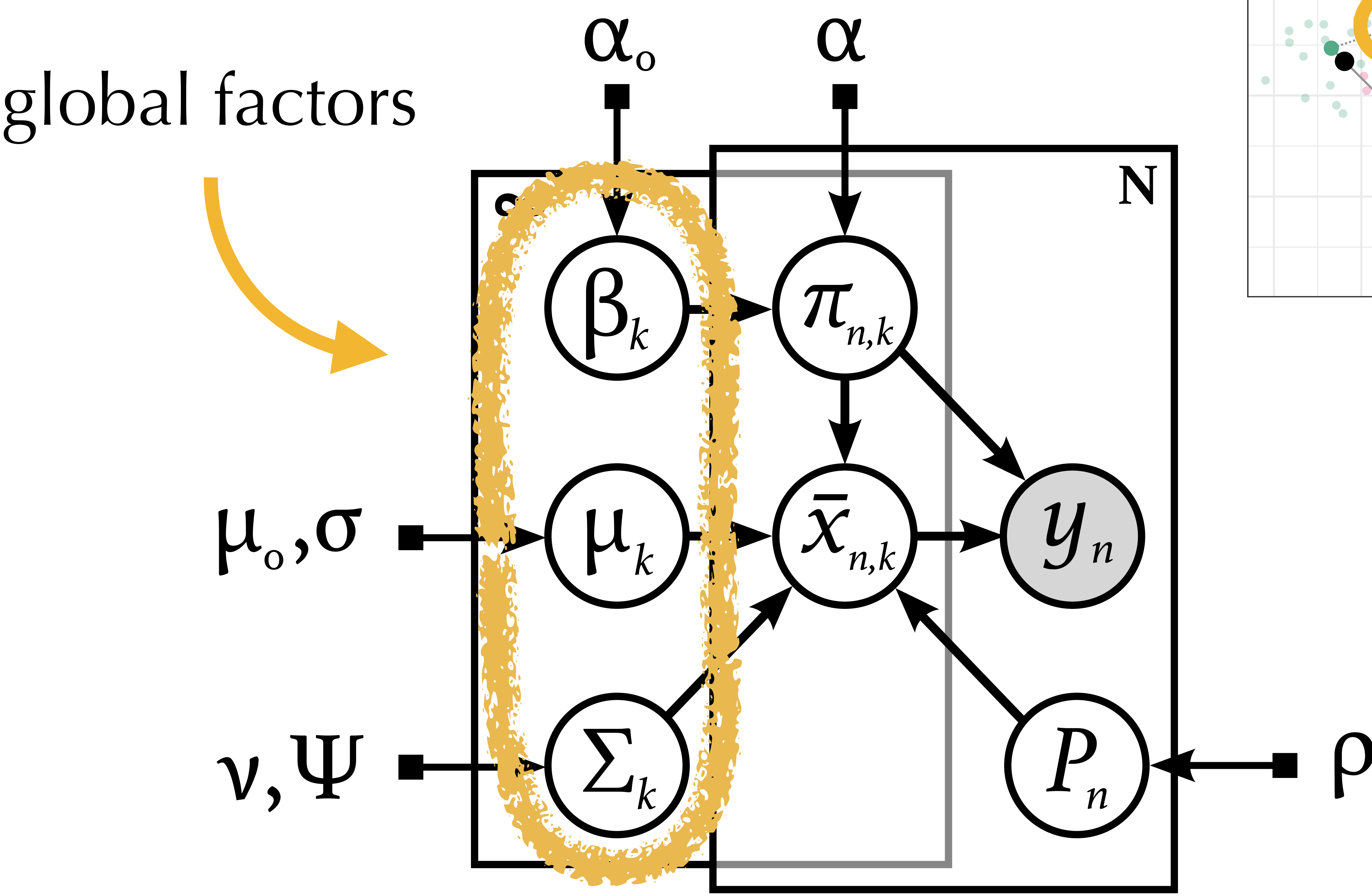


Our Model

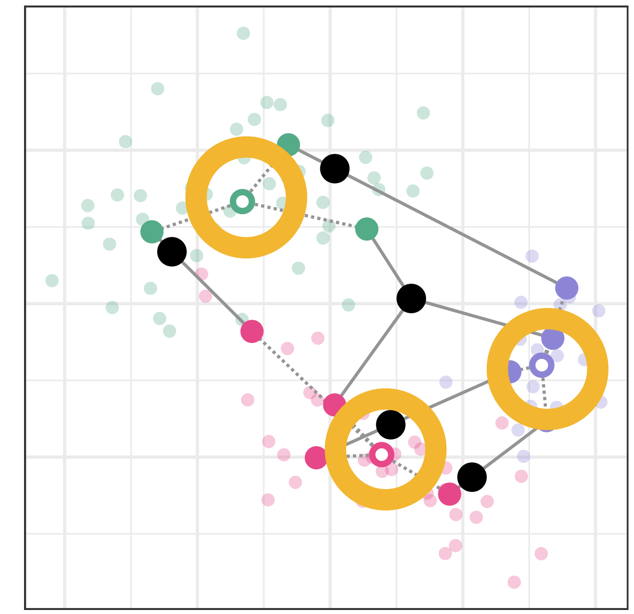


α_0
α
∞
N
β_k
$\pi_{n,k}$
μ_0,σ
μ_k
$\bar{x}_{n,k}$
y_n
ν,Ψ
Σ_k
P_n
ρ

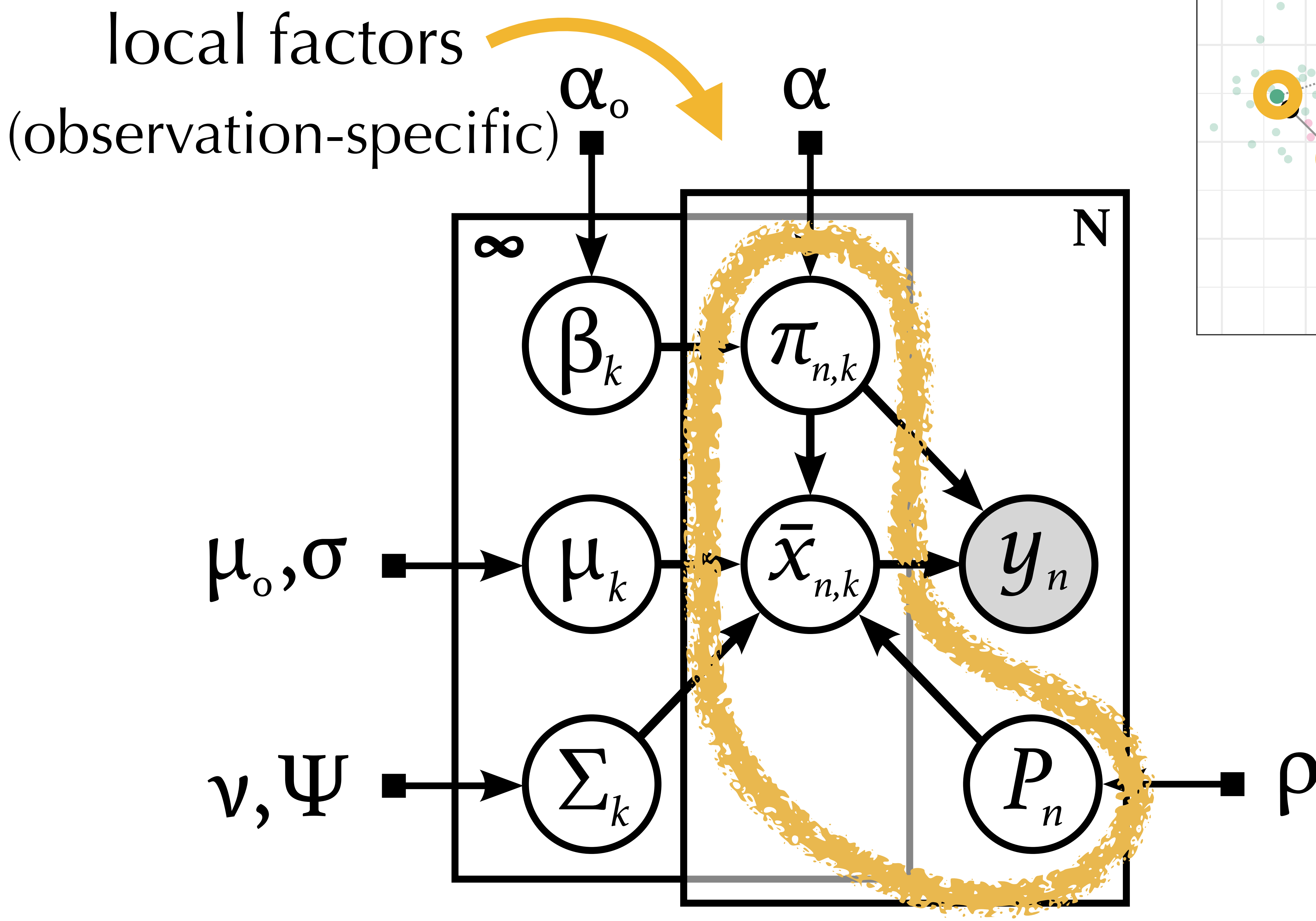
Our Model



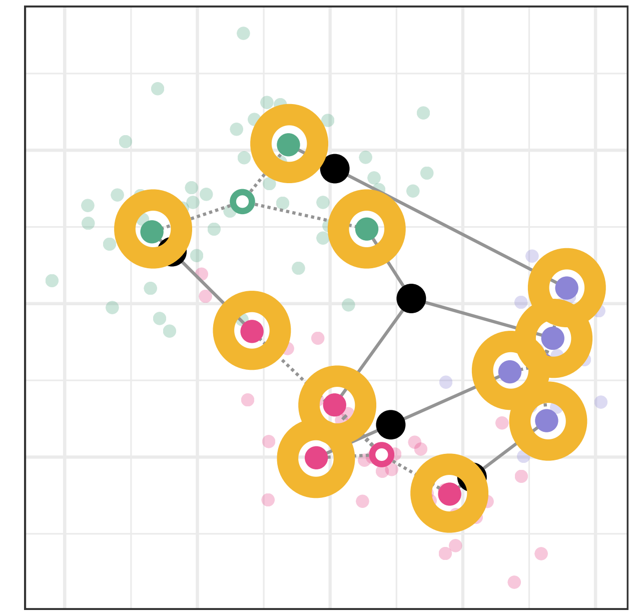
global factors
α₀
α
β_k
$\pi_{n,k}$
N
μ_0,σ
μ_k
$\bar{x}_{n,k}$
y_n
ν,Ψ
Σ_k
P_n
ρ



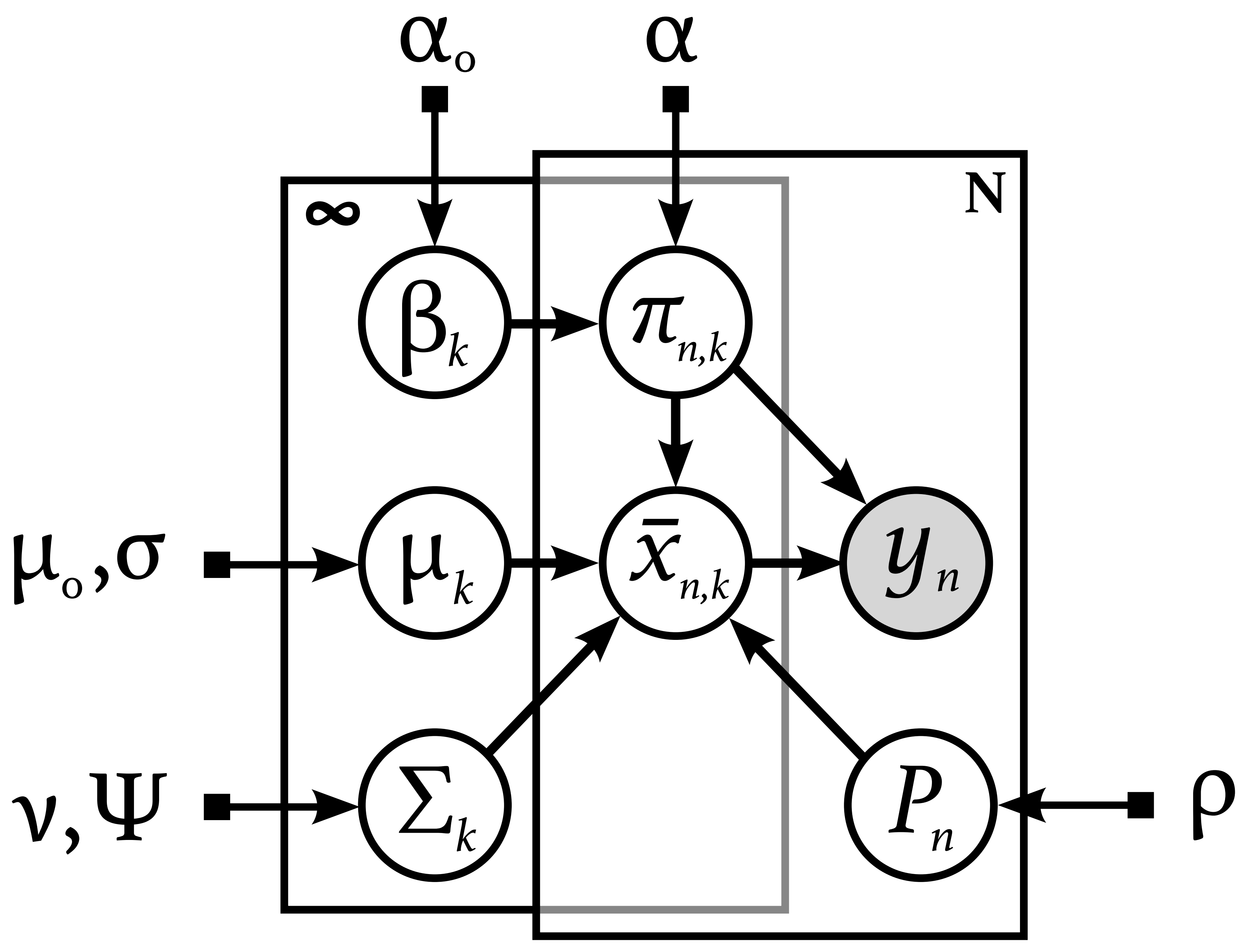
Our Model



local factors
(observation-specific)
α₀
α
∞
N
β_k
π_n,k
μ₀,σ
μ_k
x̄_n,k
y_n
ν,Ψ
Σ_k
P_n
ρ

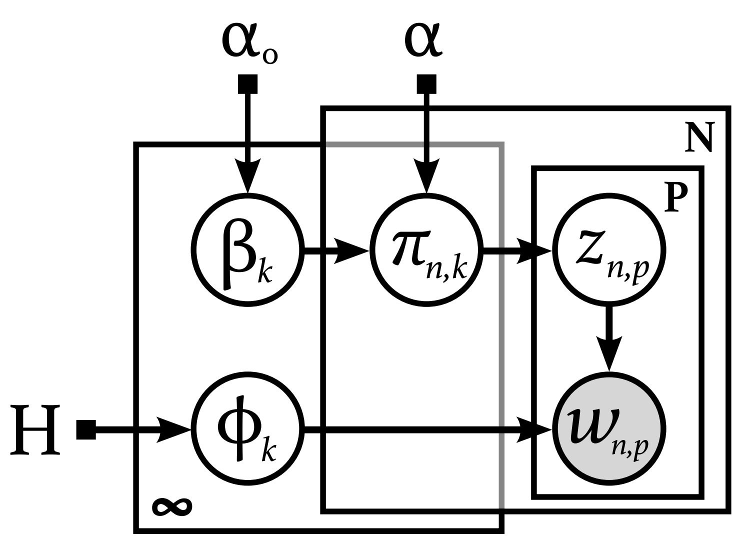


Our Model

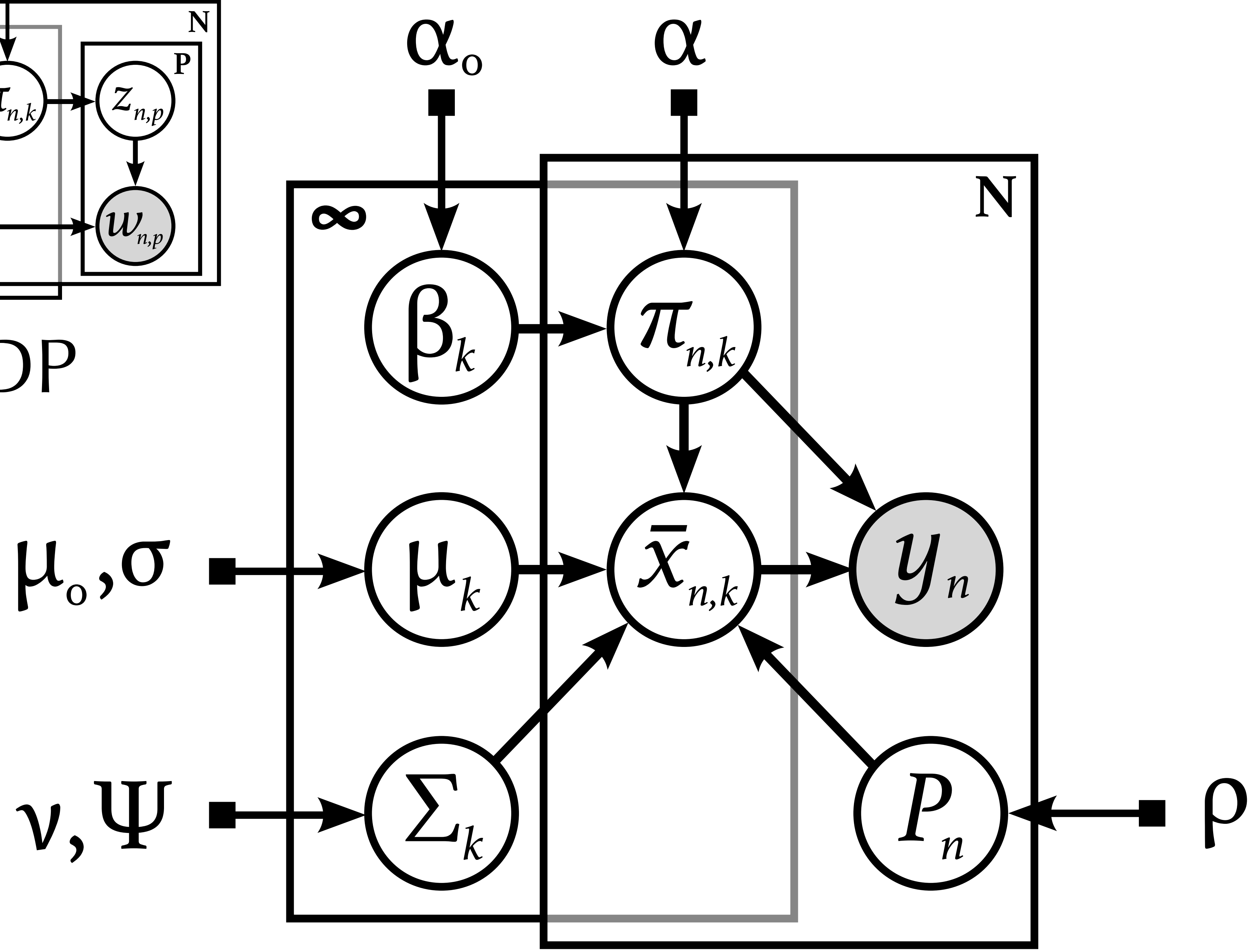


α_0
α
∞
N
β_k
$\pi_{n,k}$
μ_0, σ
μ_k
$\bar{x}_{n,k}$
y_n
ν, Ψ
Σ_k
P_n
ρ

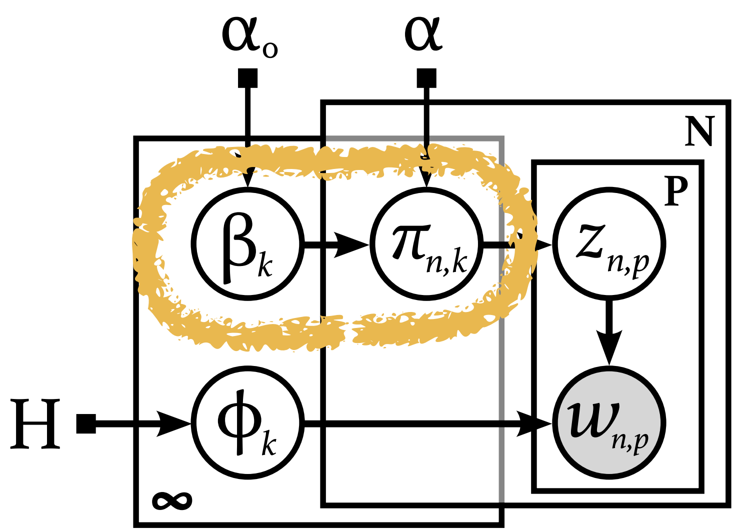
Our Model



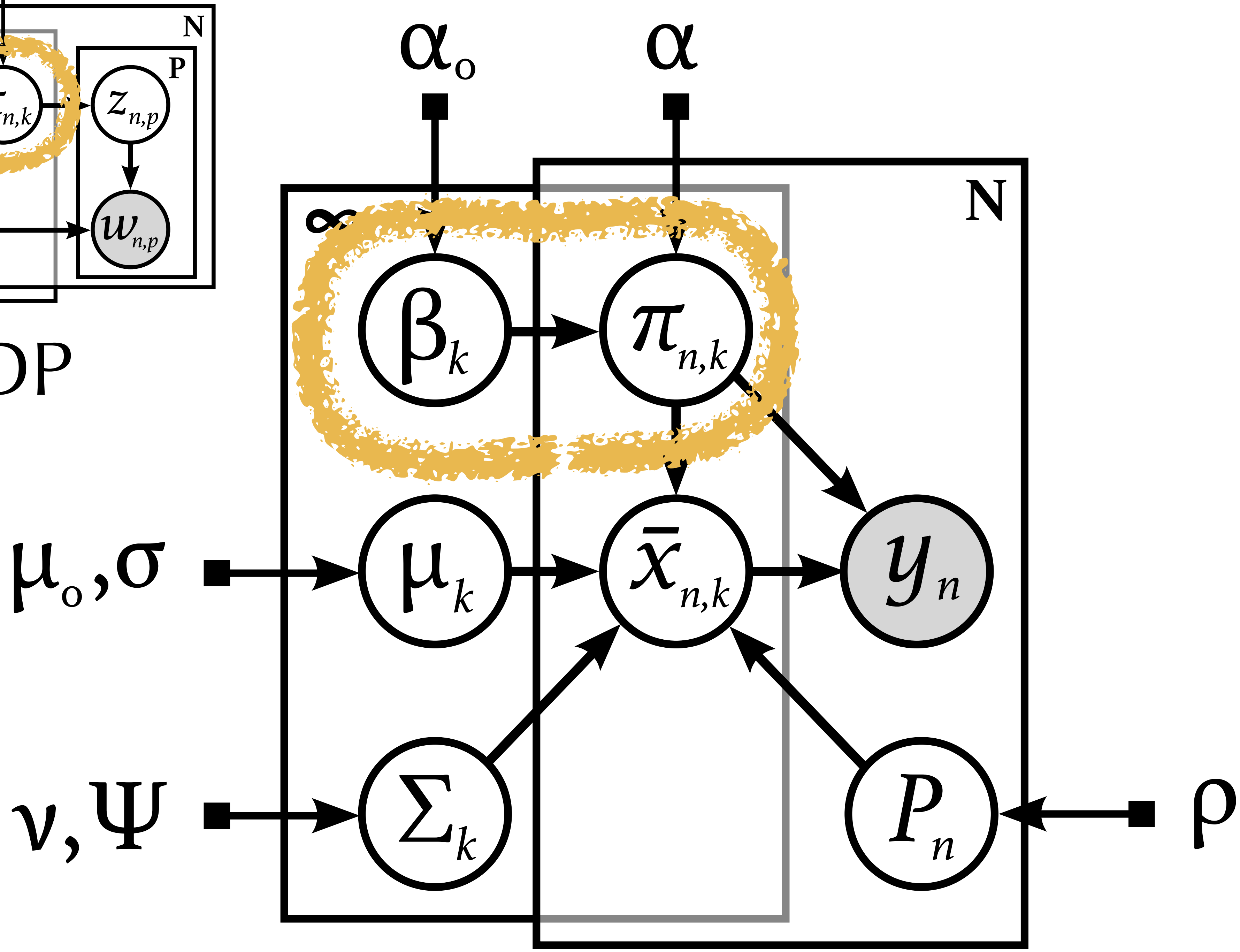
HDP



Our Model



HDP

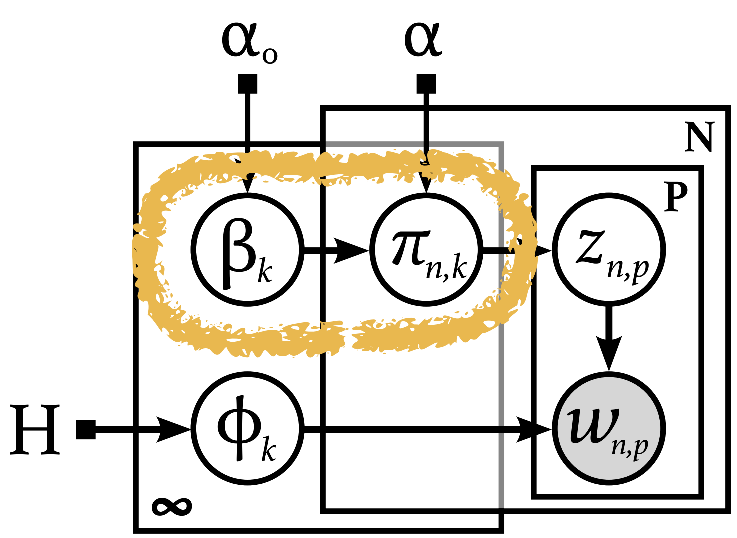


Our Model

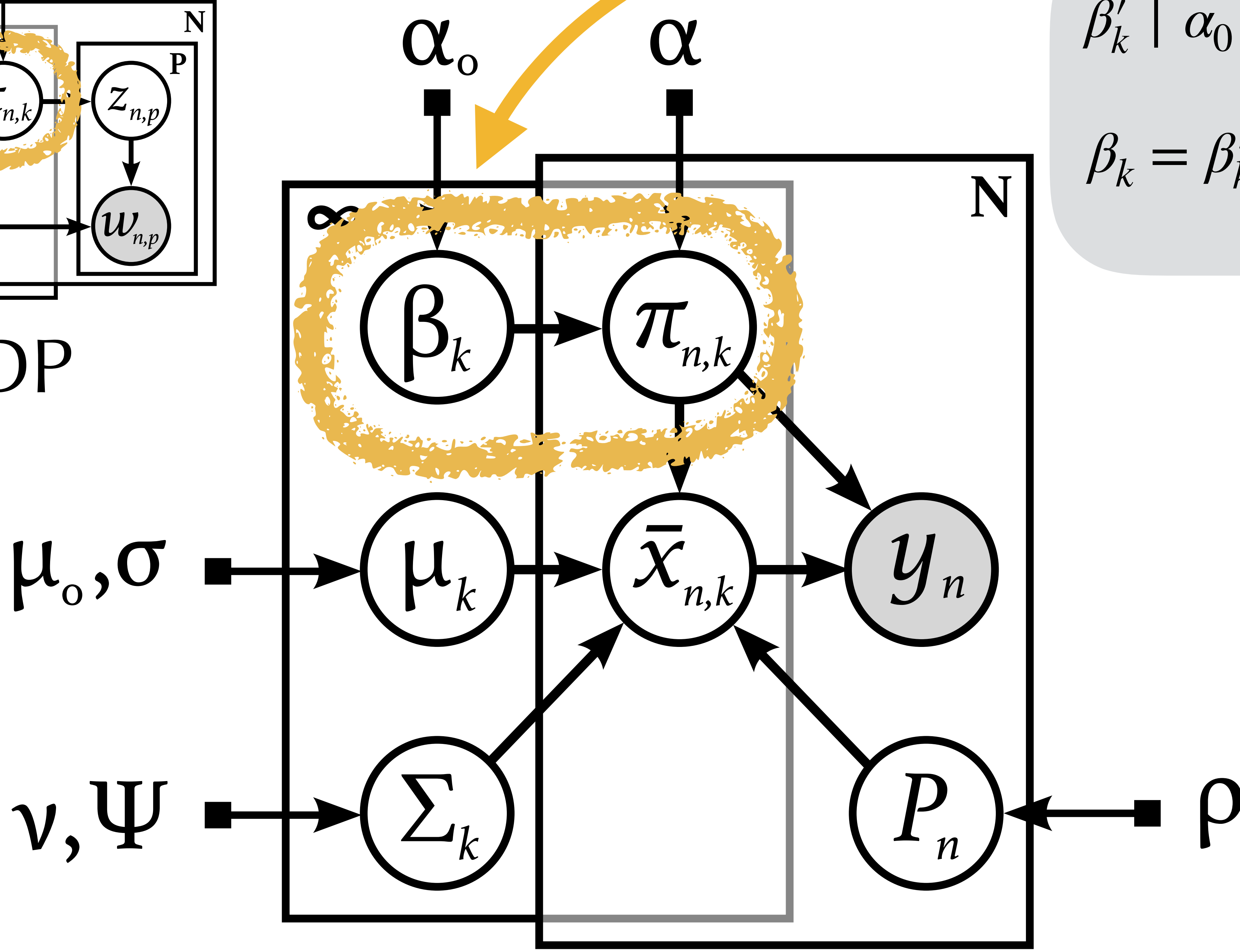
Paisley, 2012

$$\beta'_k \mid \alpha_0 \sim \text{Beta}(1,\alpha_0)$$

$$\beta_k = \beta'_k \prod_{\ell=1}^{k-1} (1-\beta'_\ell)$$

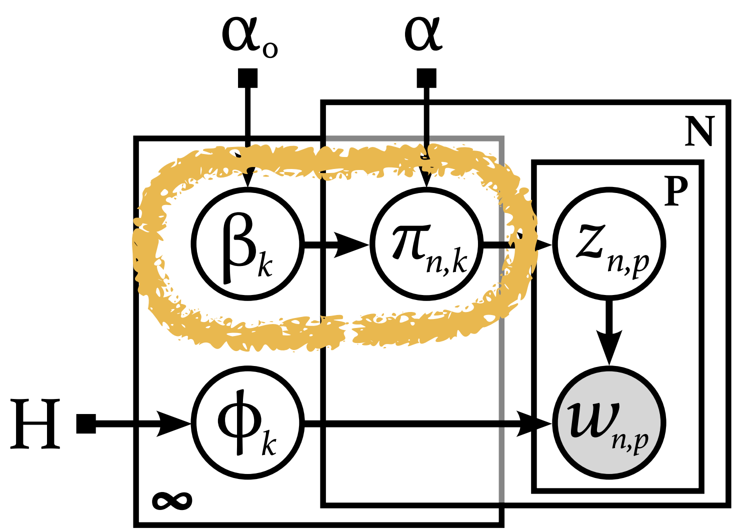


HDP

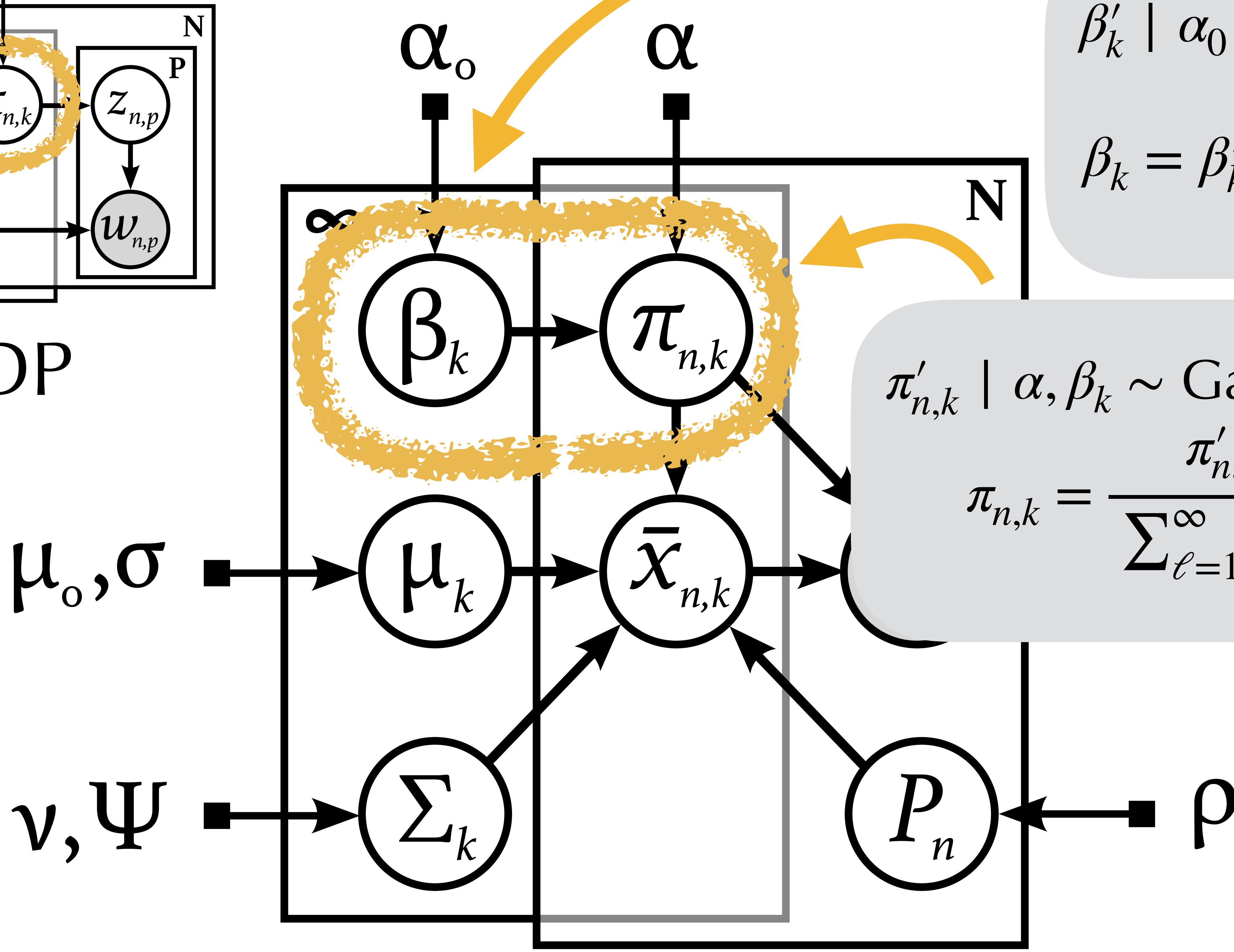


Our Model

Paisley, 2012



HDP



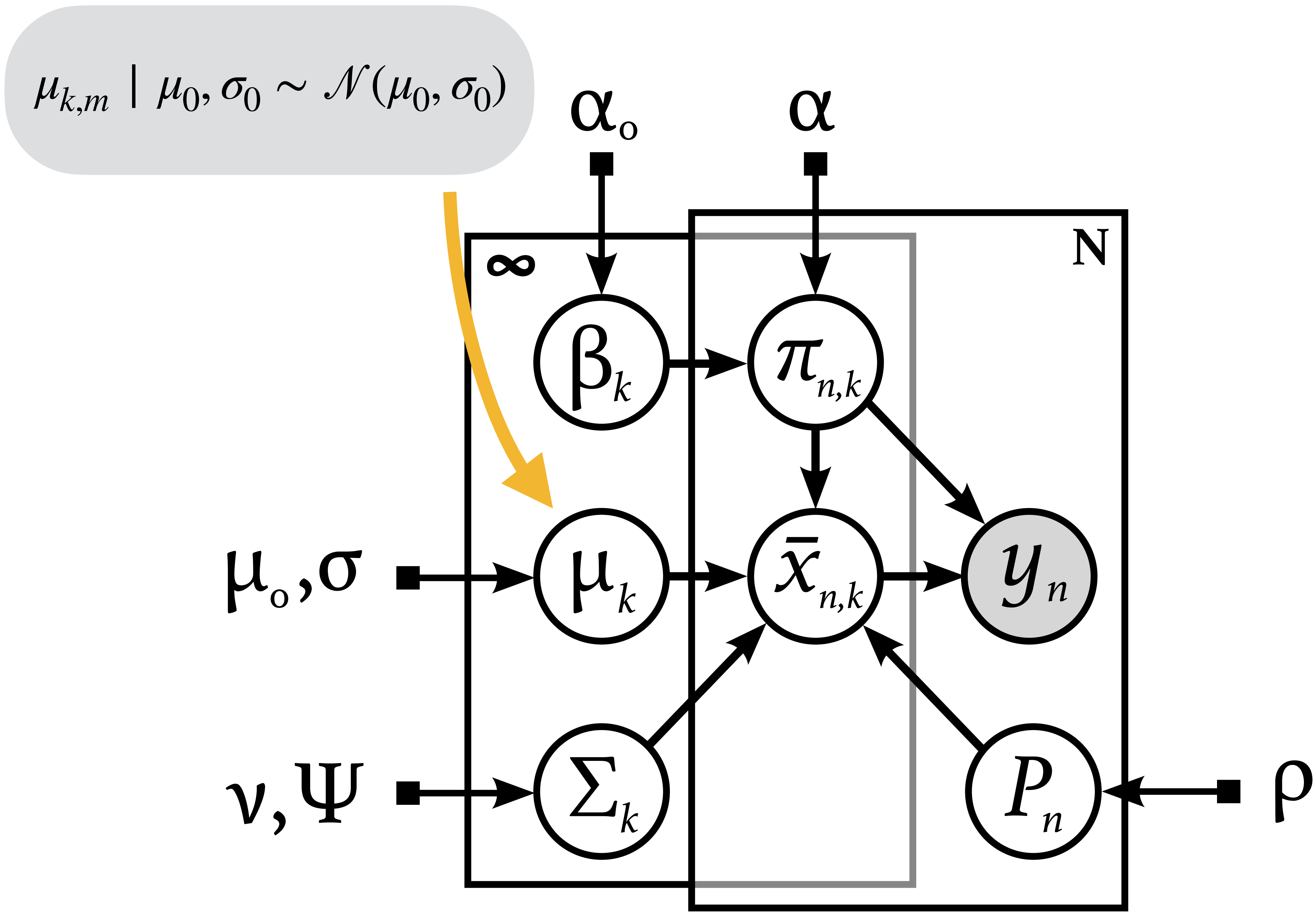
$$\beta'_k \mid \alpha_0 \sim \text{Beta}(1, \alpha_0)$$

$$\beta_k = \beta'_k \prod_{\ell=1}^{k-1} \left(1 - \beta'_\ell\right)$$

$$\pi'_{n,k} \mid \alpha, \beta_k \sim \text{Gamma}(\alpha\beta_k, 1)$$

$$\pi_{n,k} = \frac{\pi'_{n,k}}{\sum_{\ell=1}^{\infty} \pi'_{n,\ell}}$$

Our Model

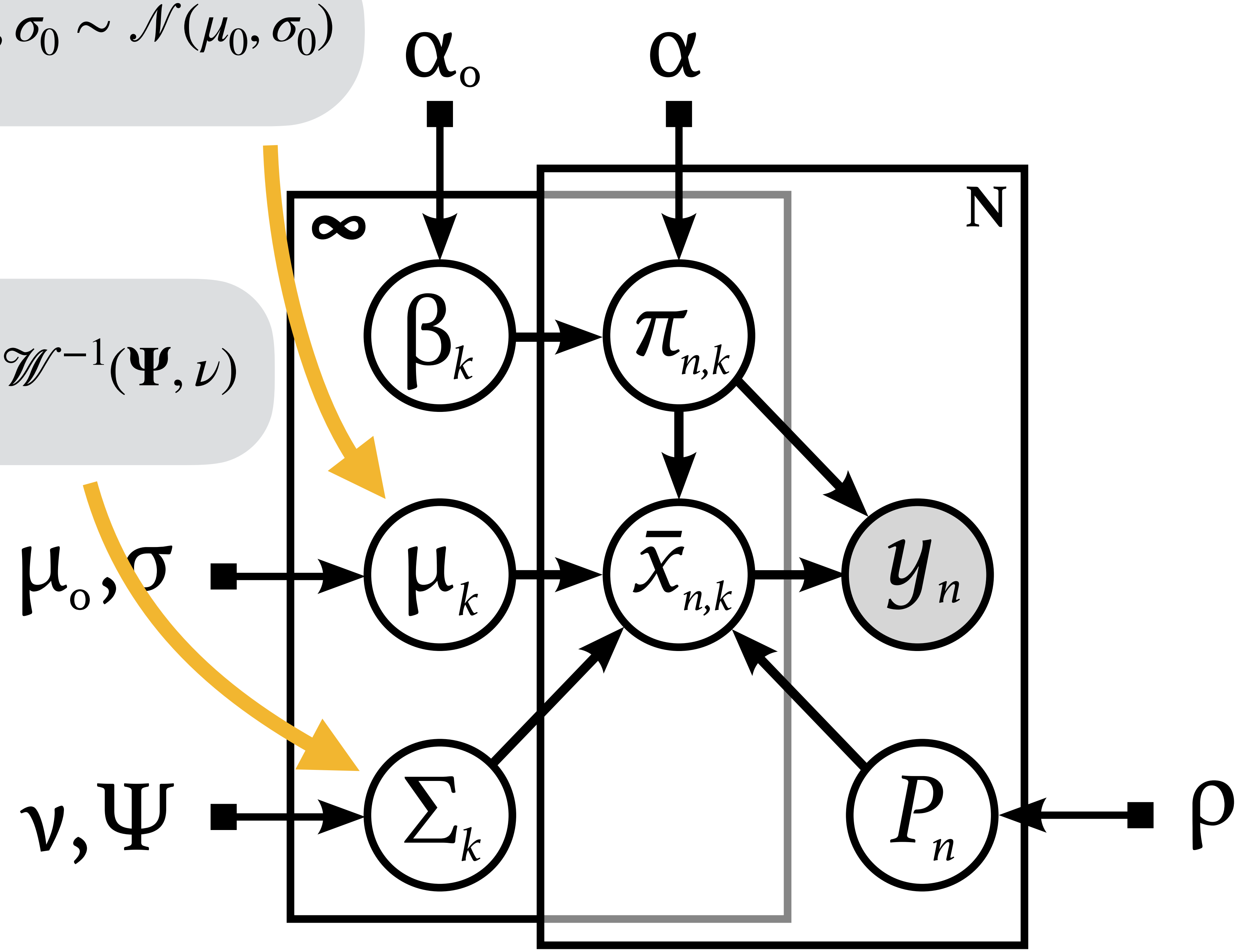


$\mu_{k,m} \mid \mu_0, \sigma_0 \sim \mathcal{N}(\mu_0, \sigma_0)$
α_o
α
∞
N
β_k
$\pi_{n,k}$
μ_o, σ
μ_k
$\bar{x}_{n,k}$
y_n
ν, Ψ
Σ_k
P_n
ρ

Our Model

$\mu_{k,m} \mid \mu_0, \sigma_0 \sim \mathcal{N}(\mu_0, \sigma_0)$

$\boldsymbol{\Sigma}_k \mid \nu, \boldsymbol{\Psi} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$

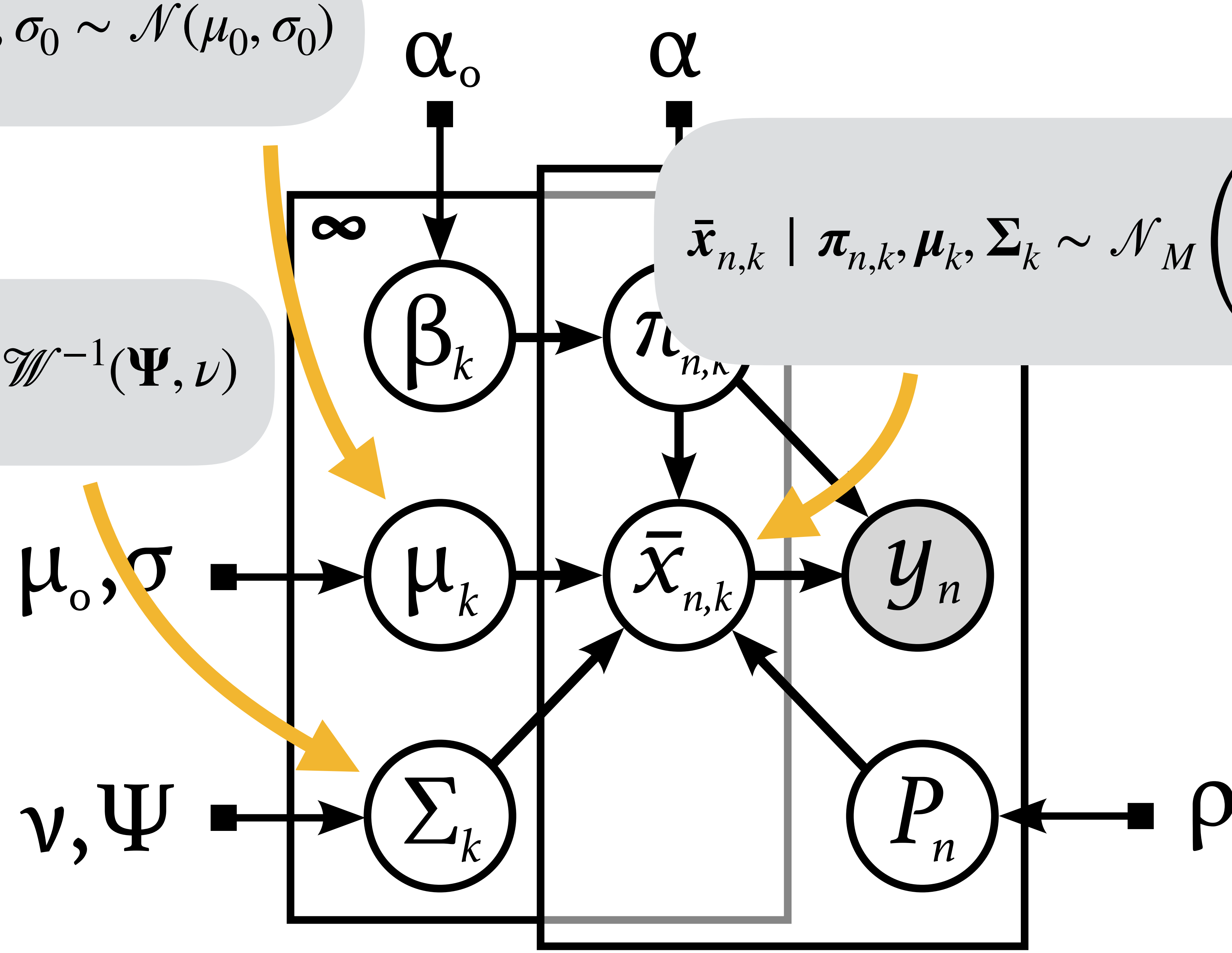


Our Model

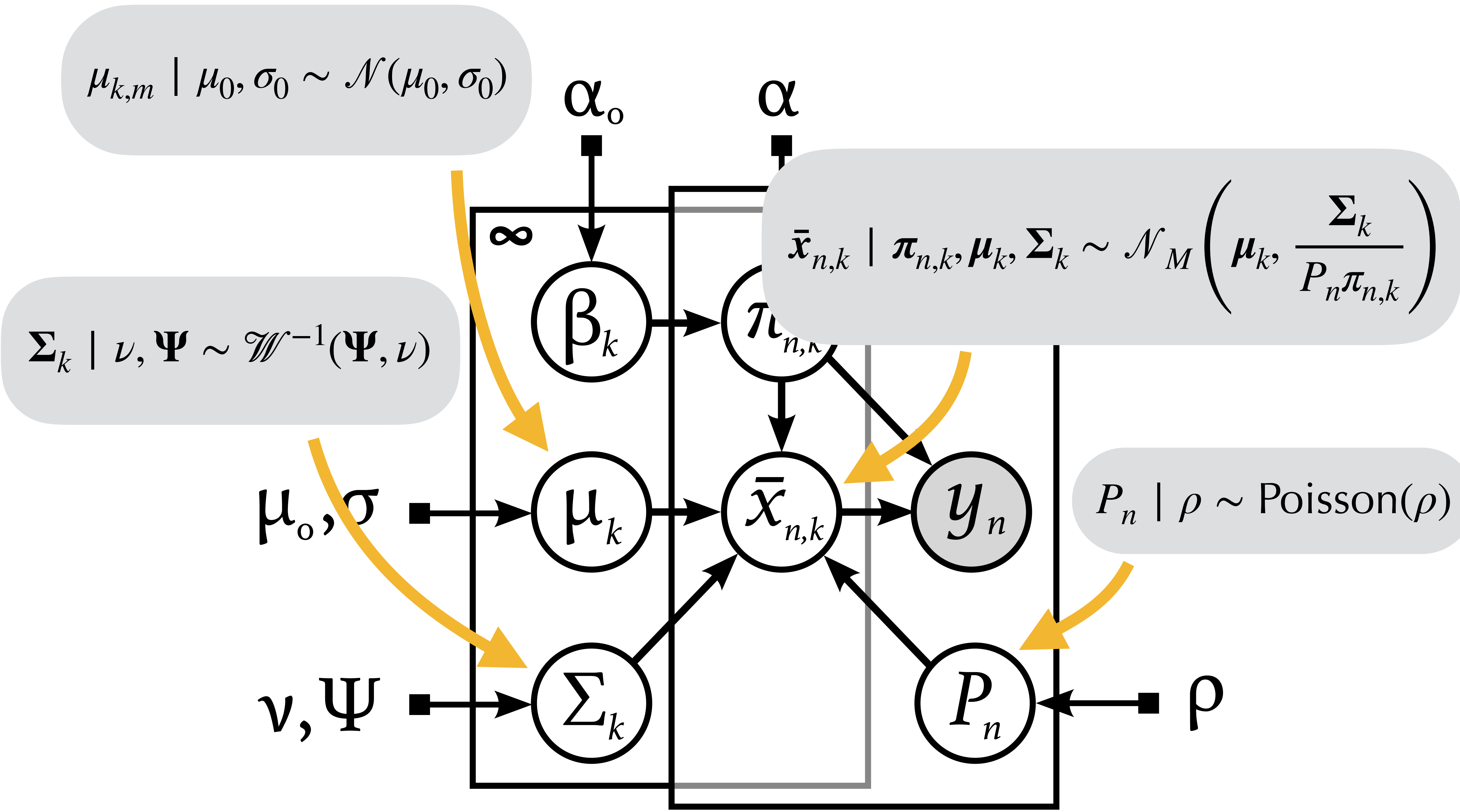
$$\mu_{k,m} \mid \mu_0, \sigma_0 \sim \mathcal{N}(\mu_0, \sigma_0)$$

$$\boldsymbol{\Sigma}_k \mid \nu, \boldsymbol{\Psi} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$$

$$\bar{\boldsymbol{x}}_{n,k} \mid \boldsymbol{\pi}_{n,k}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \sim \mathcal{N}_M\left(\boldsymbol{\mu}_k, \frac{\boldsymbol{\Sigma}_k}{P_n \pi_{n,k}}\right)$$



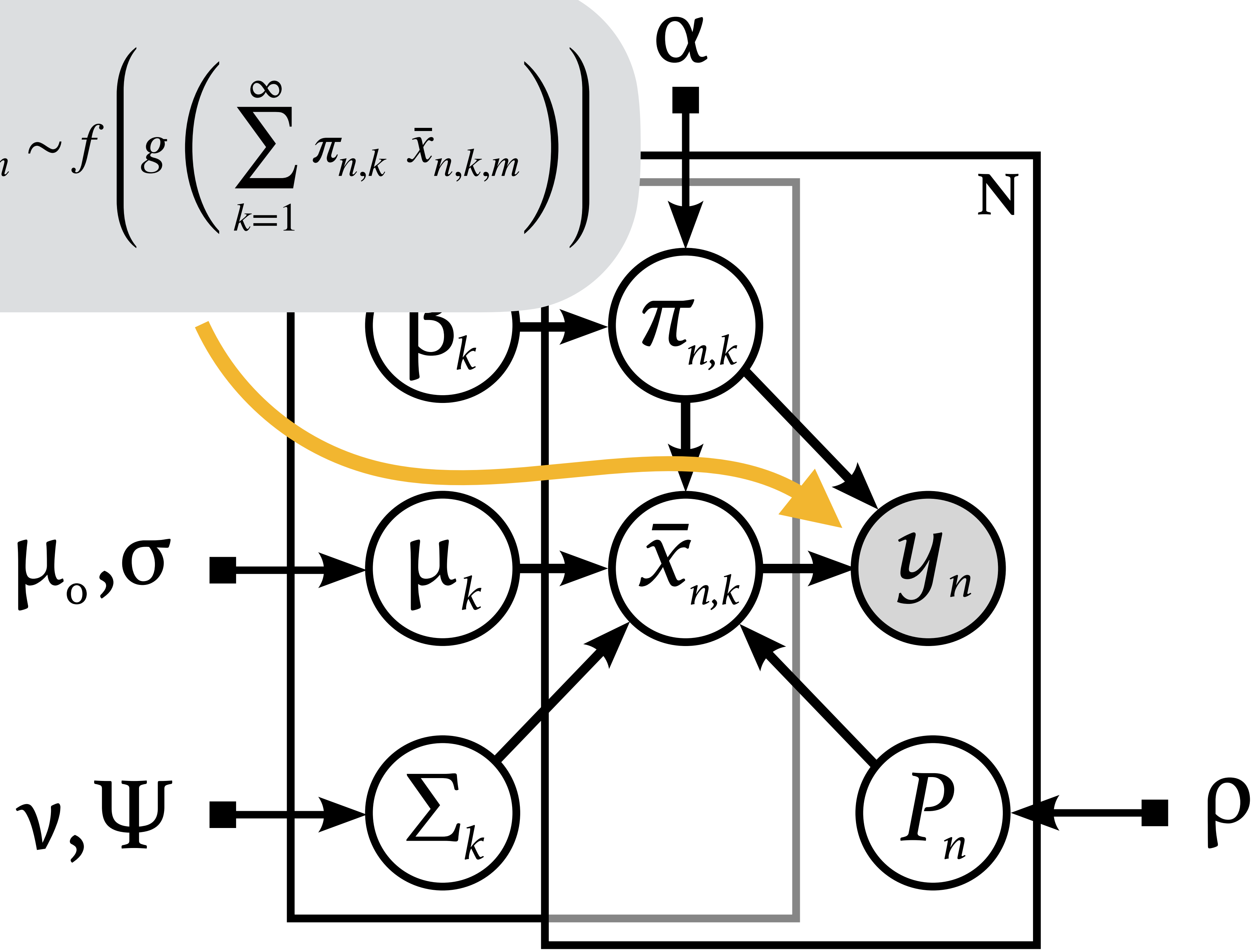
Our Model



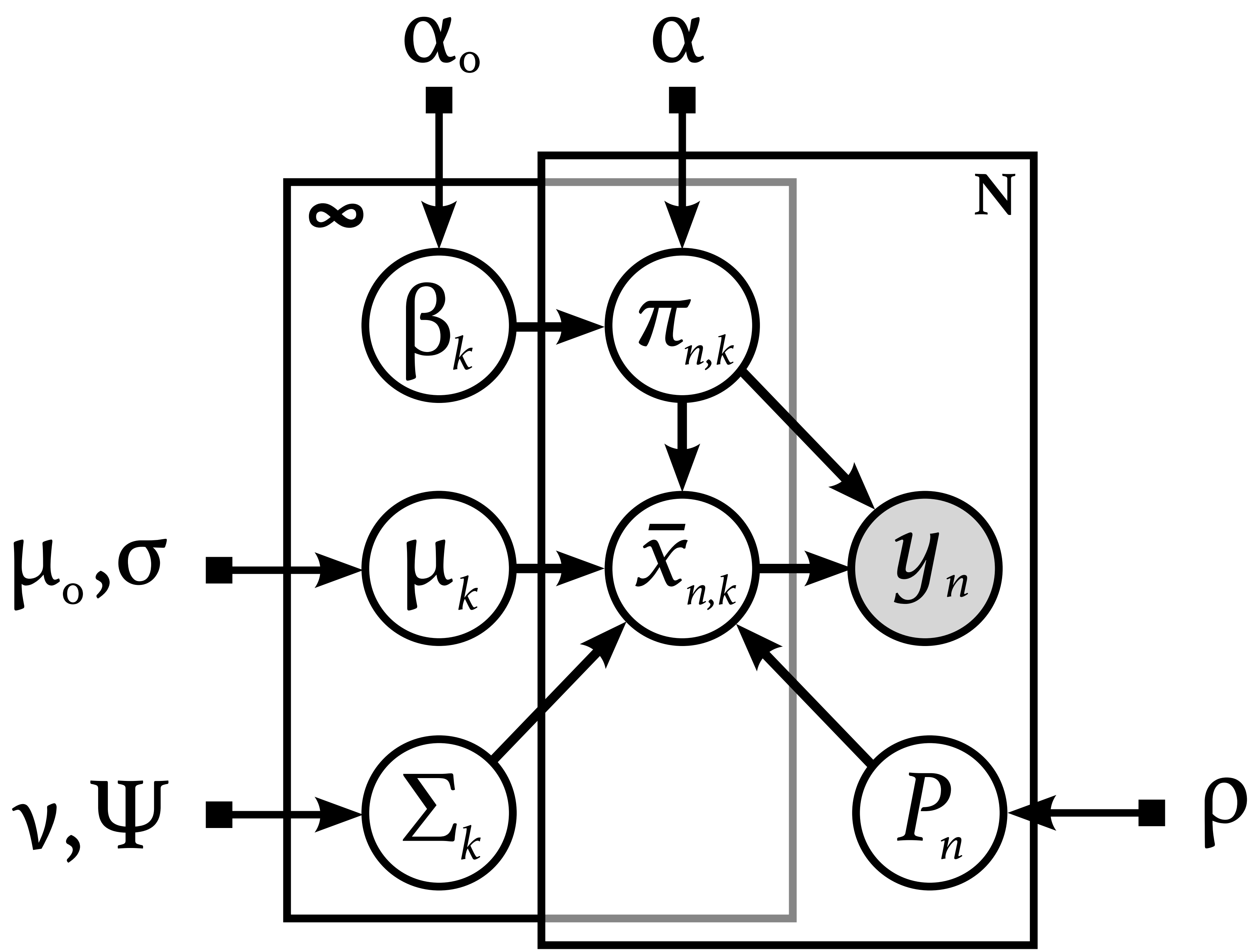
$\mu_{k,m} \mid \mu_0, \sigma_0 \sim \mathcal{N}(\mu_0, \sigma_0)$
$\boldsymbol{\Sigma}_k \mid \nu, \boldsymbol{\Psi} \sim \mathcal{W}^{-1}(\boldsymbol{\Psi}, \nu)$
$\bar{\boldsymbol{x}}_{n,k} \mid \boldsymbol{\pi}_{n,k}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \sim \mathcal{N}_M\left(\boldsymbol{\mu}_k, \frac{\boldsymbol{\Sigma}_k}{P_n \boldsymbol{\pi}_{n,k}}\right)$
$P_n \mid \rho \sim \text{Poisson}(\rho)$
α_o
α
∞
β_k
$\pi_{n,k}$
μ_o, σ
μ_k
$\bar{x}_{n,k}$
y_n
ν, Ψ
Σ_k
P_n
ρ

Our Model

$$y_{n,m} \mid \bar{\boldsymbol{x}}_n, \boldsymbol{\pi}_n \sim f\left(g\left(\sum_{k=1}^{\infty} \pi_{n,k}\ \bar{x}_{n,k,m} \right) \right)$$

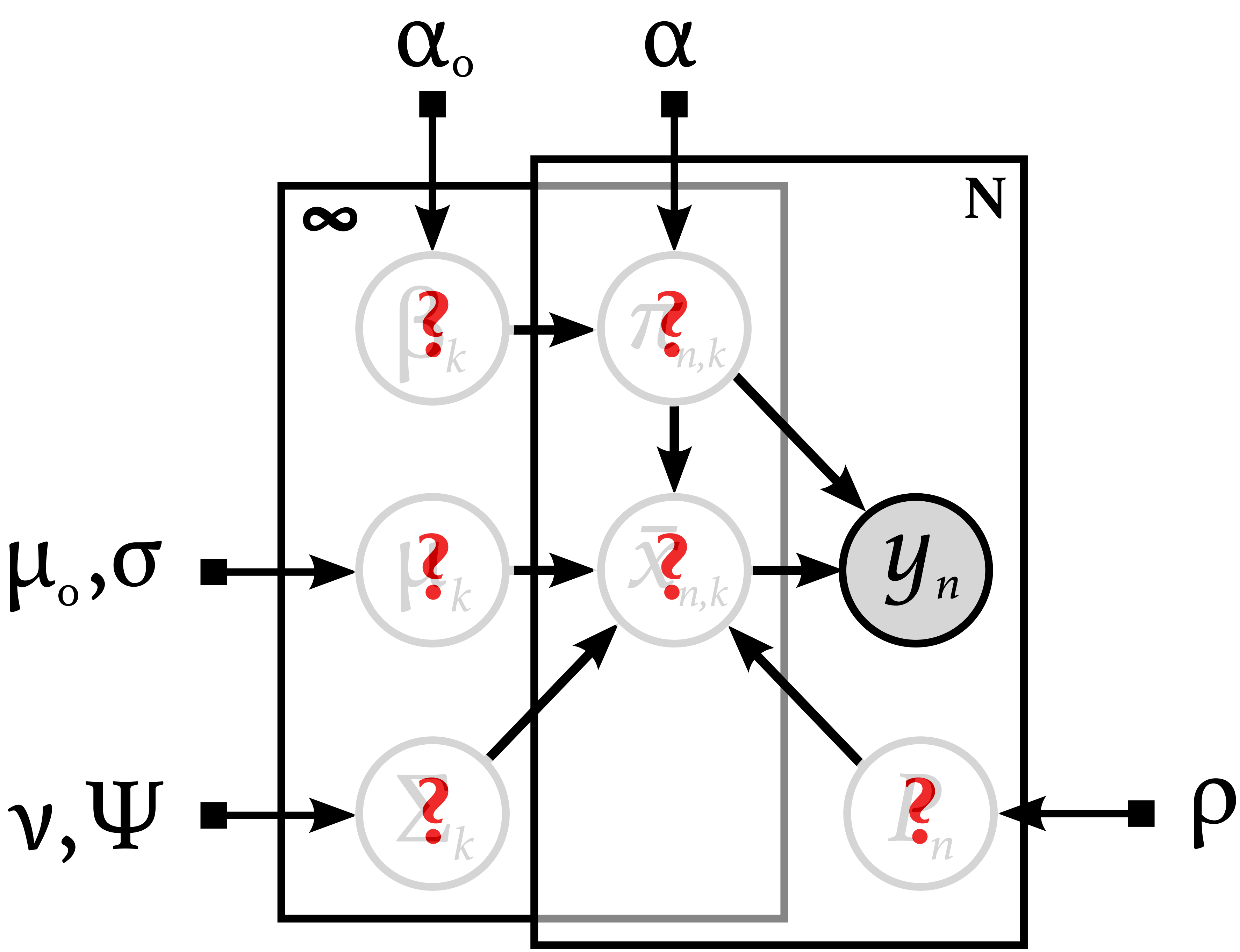


Inference



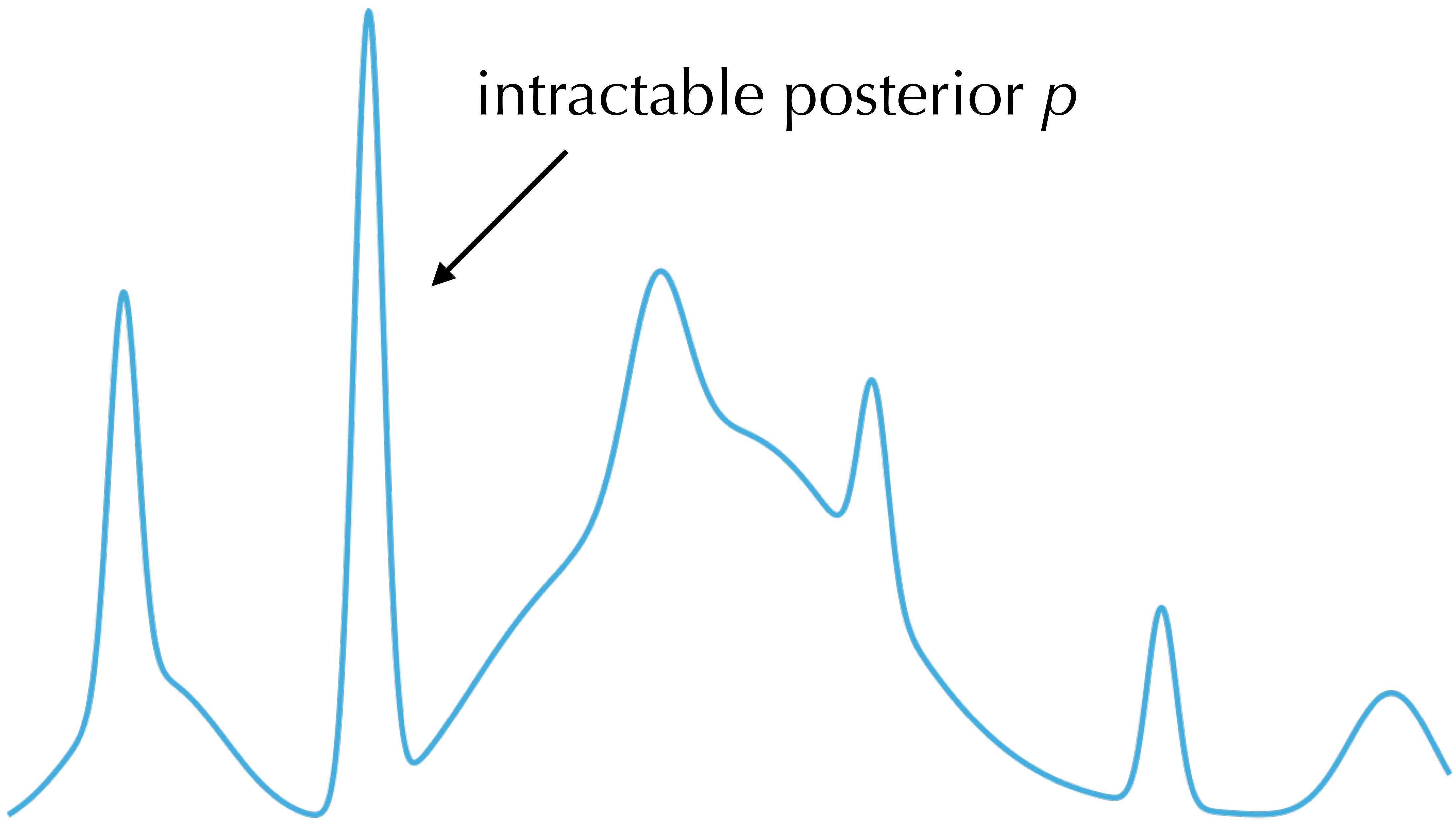
α_0
α
∞
N
β_k
$\pi_{n,k}$
μ_0, σ
μ_k
$\bar{x}_{n,k}$
y_n
ν, Ψ
Σ_k
P_n
ρ

Inference

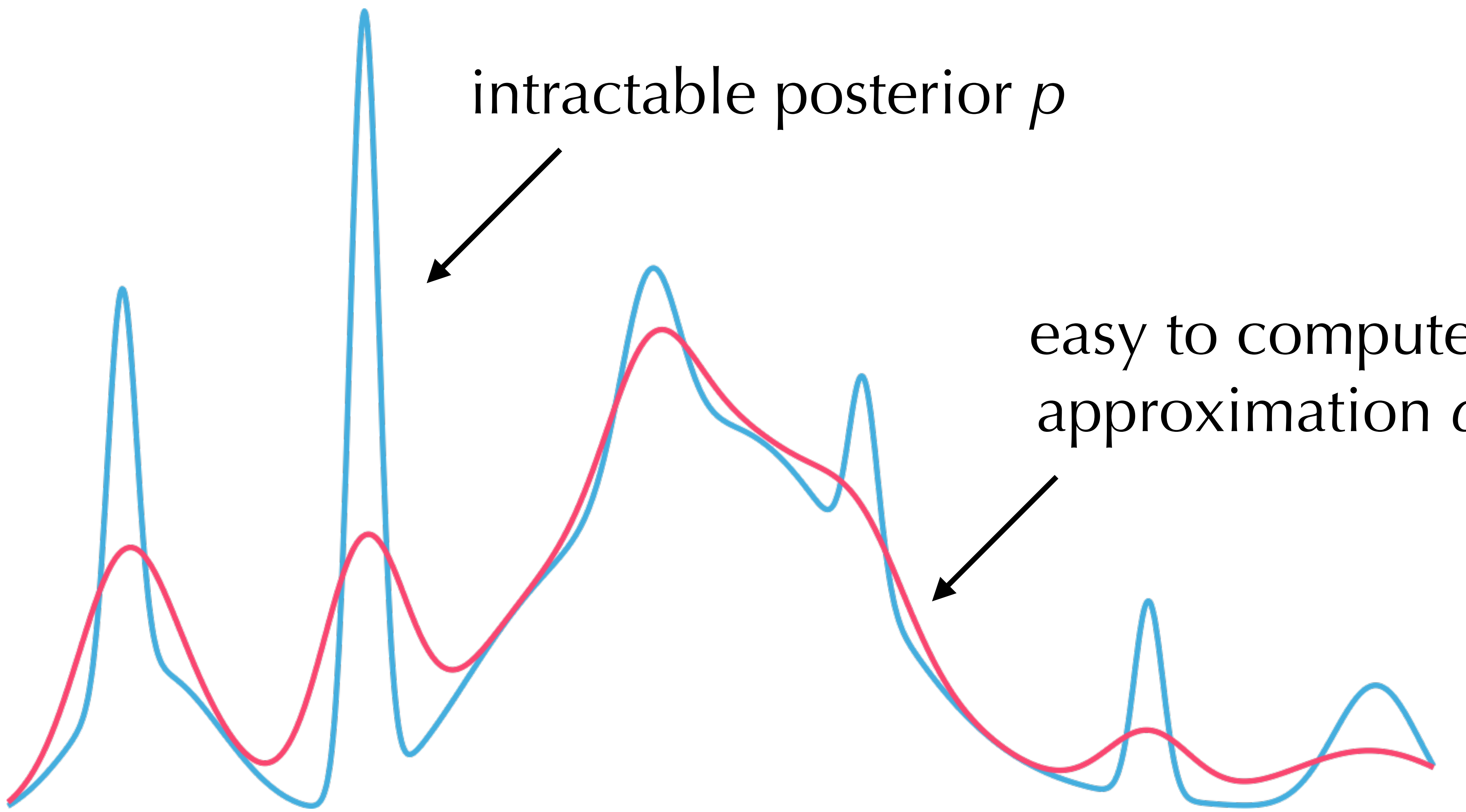


α_0
α
N
∞
β_k
$\pi_{n,k}$
μ_0, σ
μ_k
$\bar{x}_{n,k}$
y_n
ν, Ψ
Σ_k
P_n
ρ

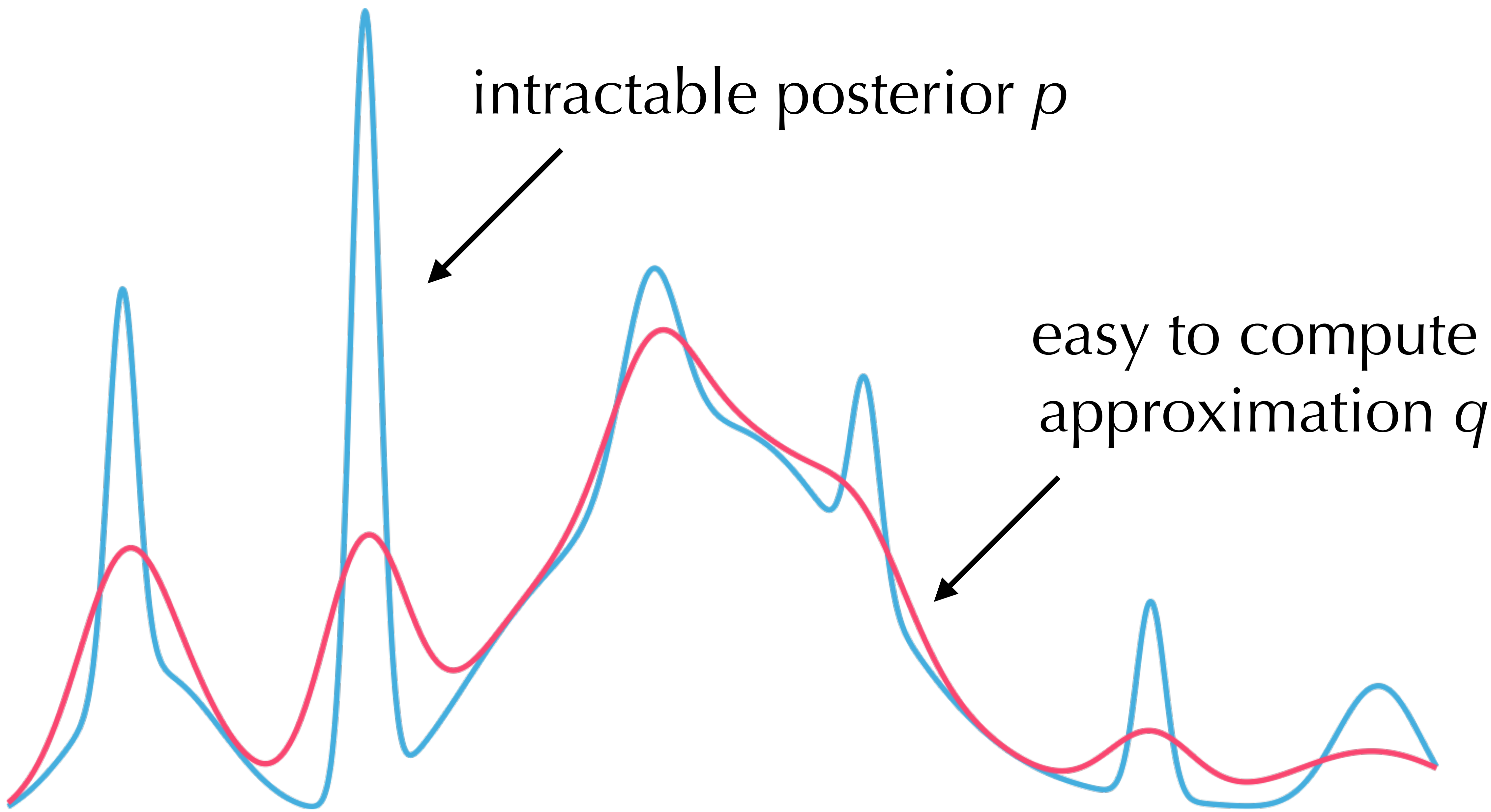
Variational Inference



Variational Inference



Variational Inference



black box variational inference (Ranganath, 2014)
split-merge procedure (Bryant, 2012) to learn K

BBVI overview

Ranganath et al., 2014

We want to estimate z

$$\nabla_{\lambda[z]}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda[z]}\log q(z \mid \boldsymbol{\lambda}[z])\left(\log p^z(\boldsymbol{y}, z, \ldots) - \log q(z \mid \boldsymbol{\lambda}[z])\right)\right]$$

BBVI overview

Ranganath et al., 2014

We want to estimate z

Which has corresponding variational parameter $\boldsymbol{\lambda}[z]$

$$\nabla_{\lambda[z]}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda[z]}\log q(z \mid \boldsymbol{\lambda}[z])\big(\log p^z(\boldsymbol{y}, z, \ldots) - \log q(z \mid \boldsymbol{\lambda}[z])\big)\right]$$

BBVI overview

Ranganath et al., 2014

$\boldsymbol{\lambda}$ is the set of all variational parameters

We want to estimate z

Which has corresponding variational parameter $\boldsymbol{\lambda}[z]$

$$\nabla_{\lambda[z]}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda[z]}\log q(z \mid \boldsymbol{\lambda}[z])\left(\log p^z(\boldsymbol{y}, z, \ldots) - \log q(z \mid \boldsymbol{\lambda}[z])\right)\right]$$

BBVI overview

Ranganath et al., 2014

$\boldsymbol{\lambda}$ is the set of all variational parameters

The gradient of the ELBO

We want to estimate z

Which has corresponding variational parameter $\boldsymbol{\lambda}[z]$

$$\nabla_{\lambda[z]}\mathcal{L} = \mathbb{E}_q\Big[\nabla_{\lambda[z]}\log q(z \mid \boldsymbol{\lambda}[z])\big(\log p^z(\boldsymbol{y}, z, \ldots) - \log q(z \mid \boldsymbol{\lambda}[z])\big)\Big]$$

BBVI overview

Ranganath et al., 2014

The gradient of the ELBO

We want to estimate z

Which has corresponding variational parameter $\boldsymbol{\lambda}[z]$

$\boldsymbol{\lambda}$ is the set of all variational parameters

$$\nabla_{\lambda[z]}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda[z]}\log q(z \mid \boldsymbol{\lambda}[z])\big(\log p^z(\mathbf{y}, z, \ldots) - \log q(z \mid \boldsymbol{\lambda}[z])\big)\right]$$

$$\approx \tilde{\nabla}_{\lambda[z]}\mathscr{L}$$

If we can **approximate this gradient**, we can use **standard stochastic gradient ascent** to update $\boldsymbol{\lambda}[z]$

BBVI overview

Ranganath et al., 2014

$\boldsymbol{\lambda}$ is the set of all variational parameters

The gradient of the ELBO

We want to estimate z

Which has corresponding variational parameter $\boldsymbol{\lambda}[z]$

$$\nabla_{\lambda[z]}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda[z]}\log q(z \mid \boldsymbol{\lambda}[z])\left(\log p^z(\boldsymbol{y}, z, \ldots) - \log q(z \mid \boldsymbol{\lambda}[z])\right)\right]$$

$$\approx \tilde{\nabla}_{\lambda[z]}\mathscr{L}$$

If we can **approximate this gradient**, we can use **standard stochastic gradient ascent** to update $\boldsymbol{\lambda}[z]$

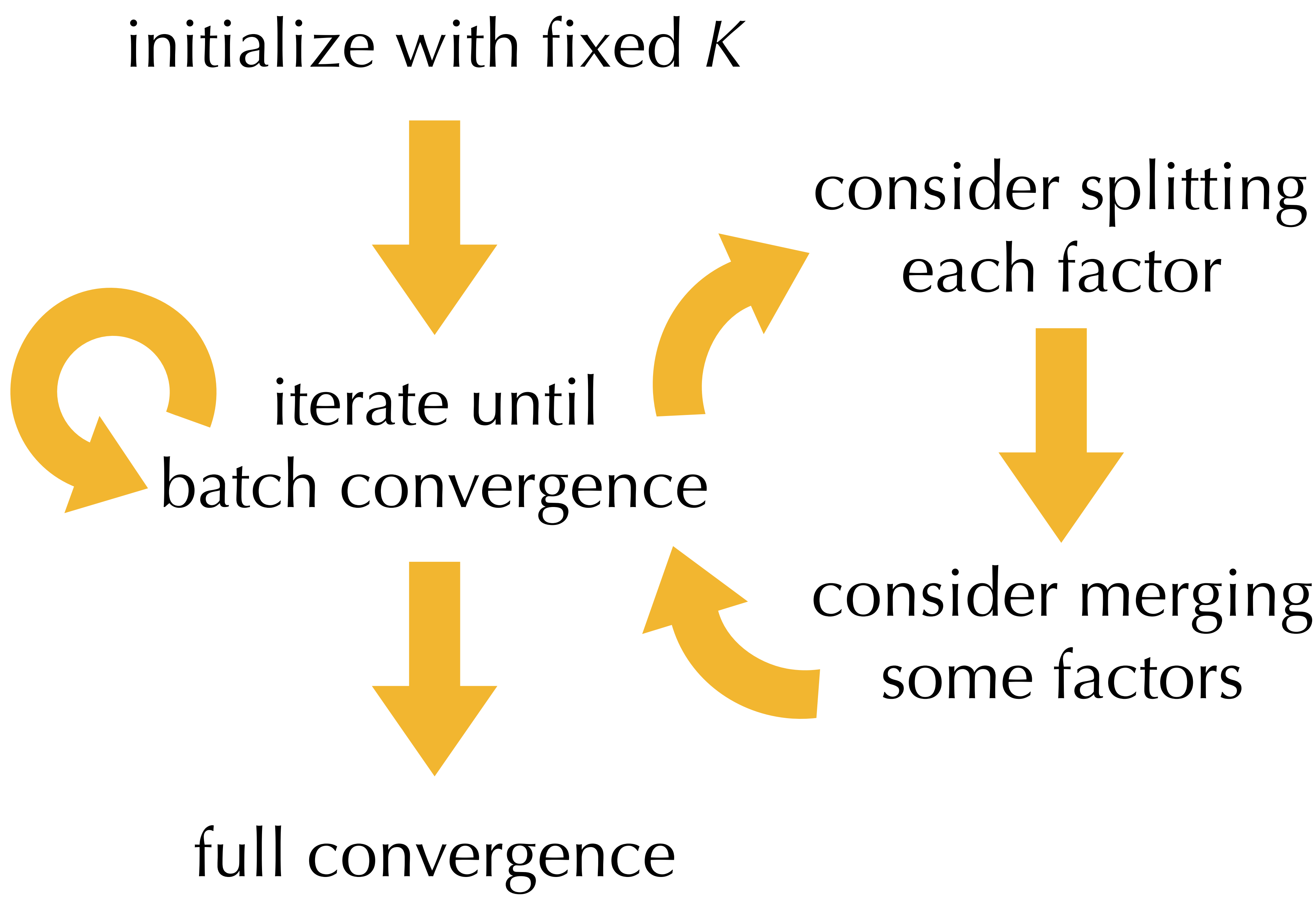
$$= \frac{1}{S}\sum_{s=1}^{S}\left[\nabla_{\lambda[z]}\log q(z[s] \mid \lambda[z])\left(\log p^z(\boldsymbol{y}, z[s], \ldots) - \log q(z[s] \mid \lambda[z])\right)\right]$$

Average over S samples

From the variational distribution $z[s] \sim q(z \mid \lambda[z])$

split/merge overview

Bryant and Sudderth, 2012



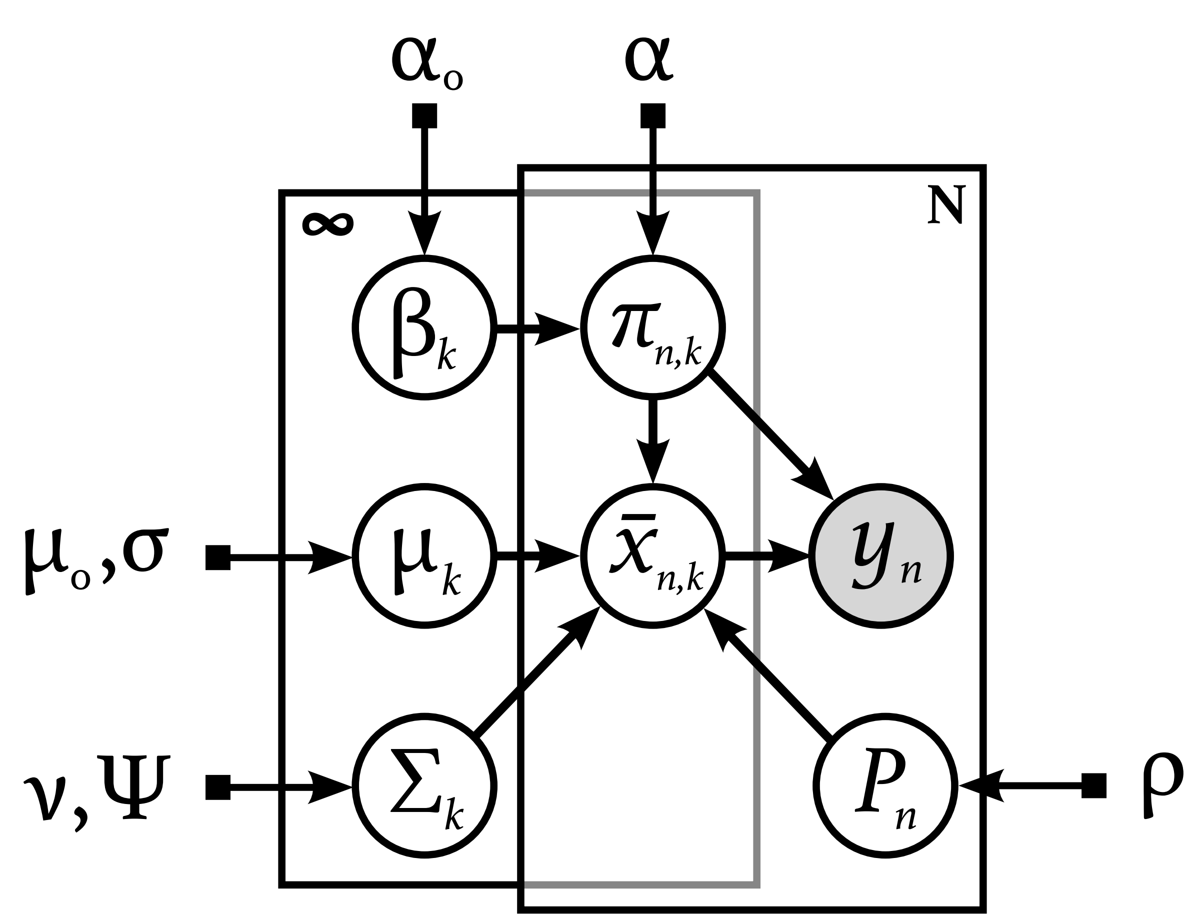
split/merge overview

Bryant and Sudderth, 2012

K

consider splitting
each factor

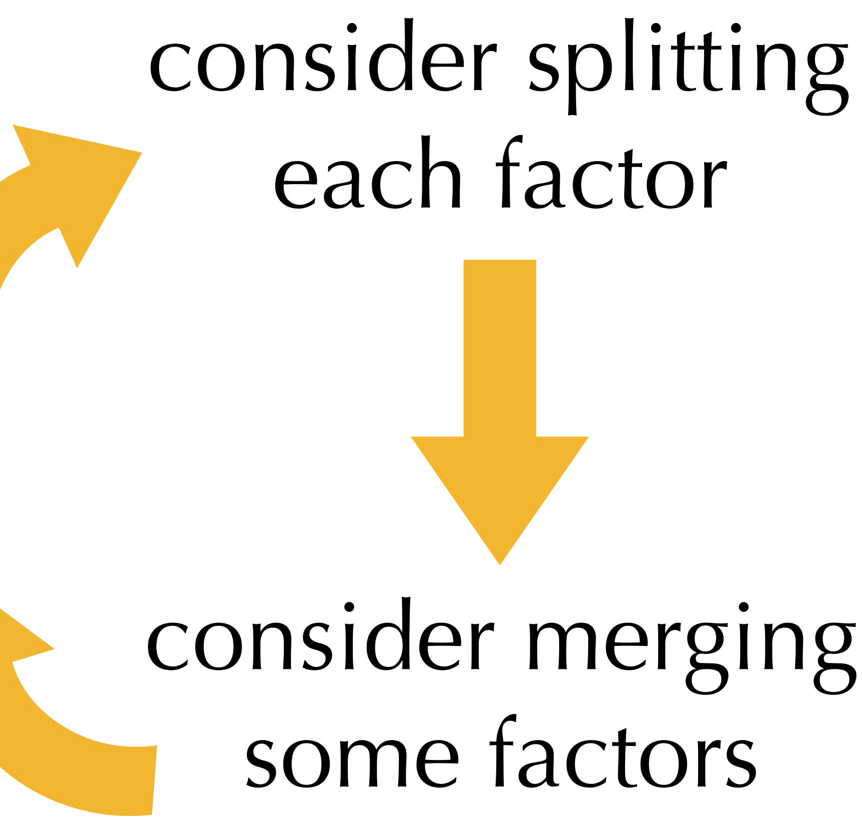
consider merging
some factors

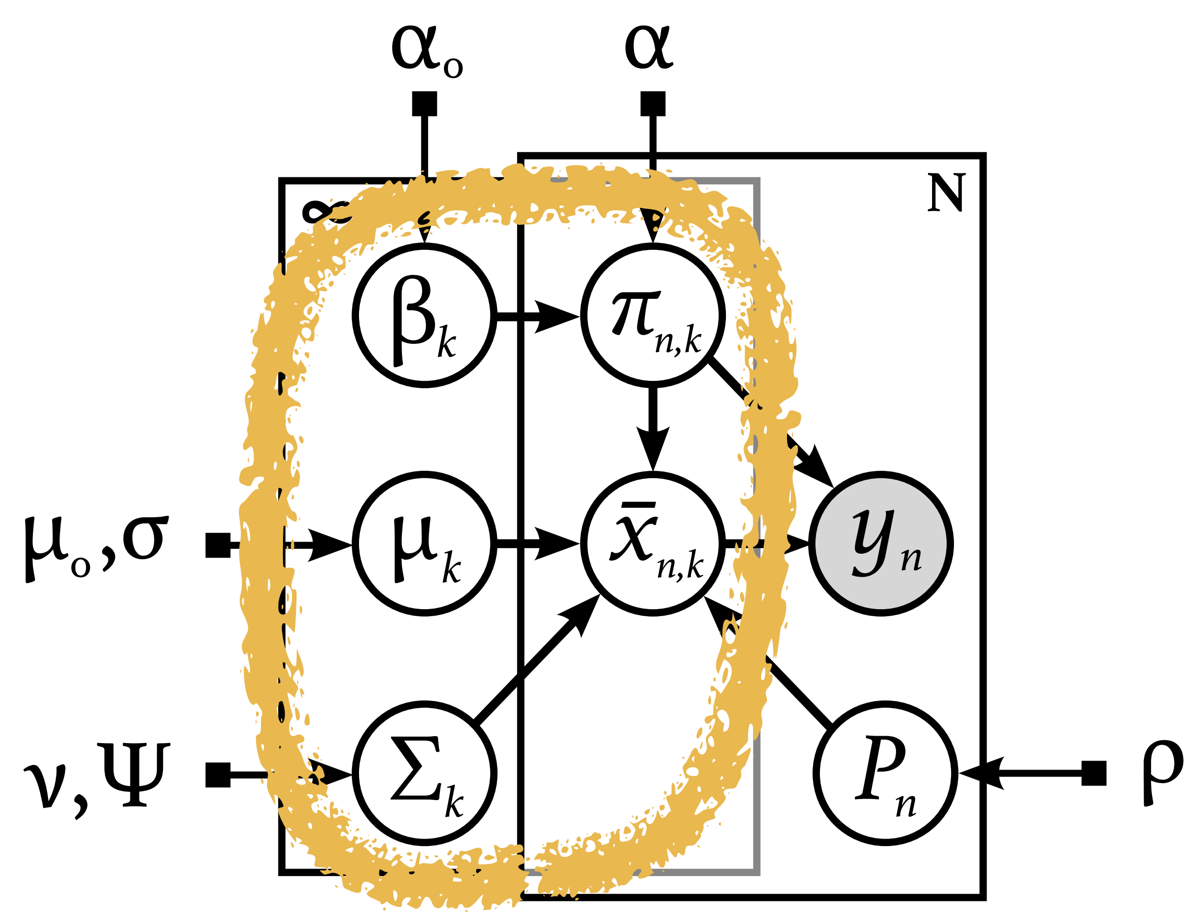


split/merge overview

Bryant and Sudderth, 2012

K





split/merge overview

Bryant and Sudderth, 2012

K

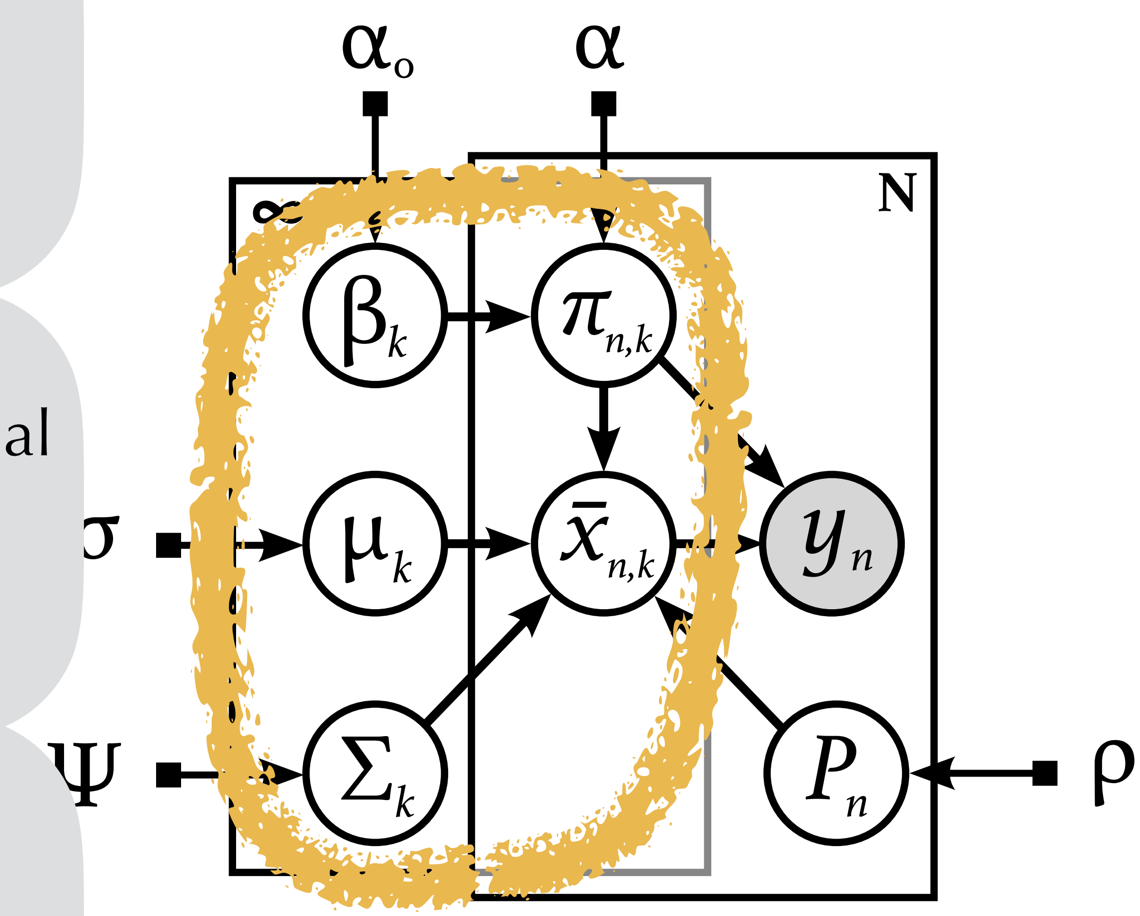
consider splitting each factor

consider merging some factors

initialize variational parameters

update variational parameters (one iteration)

accept / reject based on ELBO



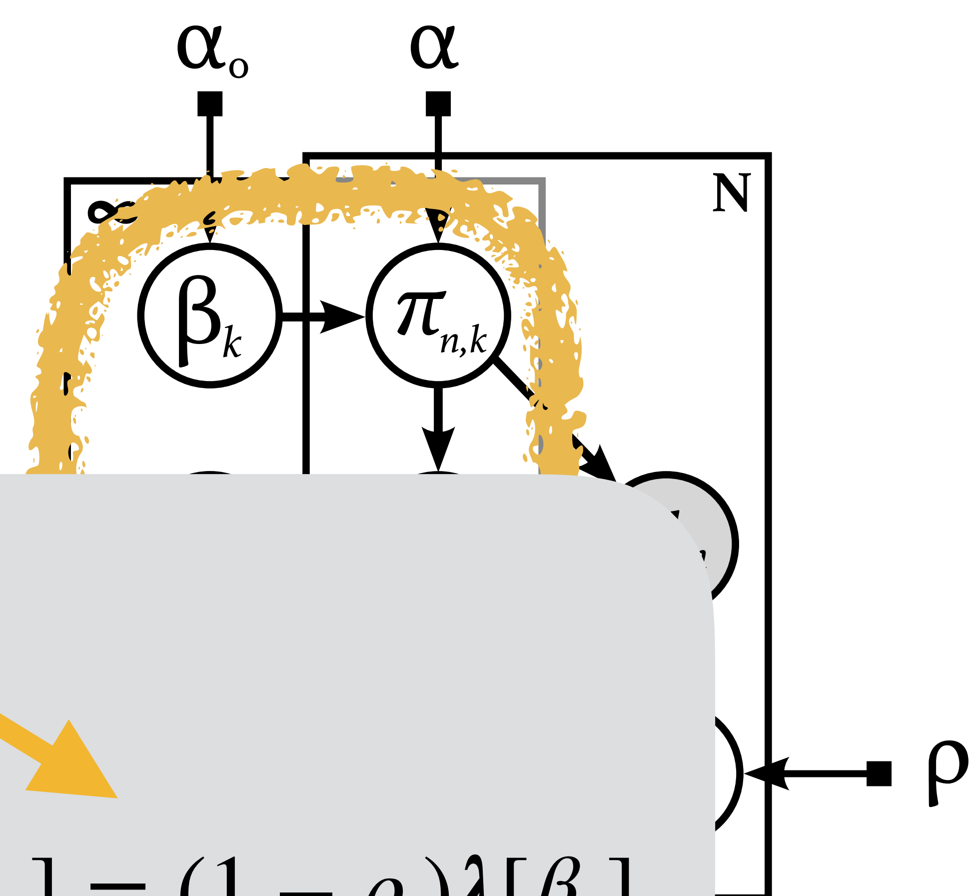
split/merge overview

Bryant and Sudderth, 2012

K

consider splitting
each factor

conside
some



$$\lambda[\beta_k]$$

$$\lambda^S[\beta_{k'}] = \rho_t \lambda[\beta_k] \qquad \lambda^S[\beta_{k''}] = (1 - \rho_t)\lambda[\beta_k]$$

split/merge overview

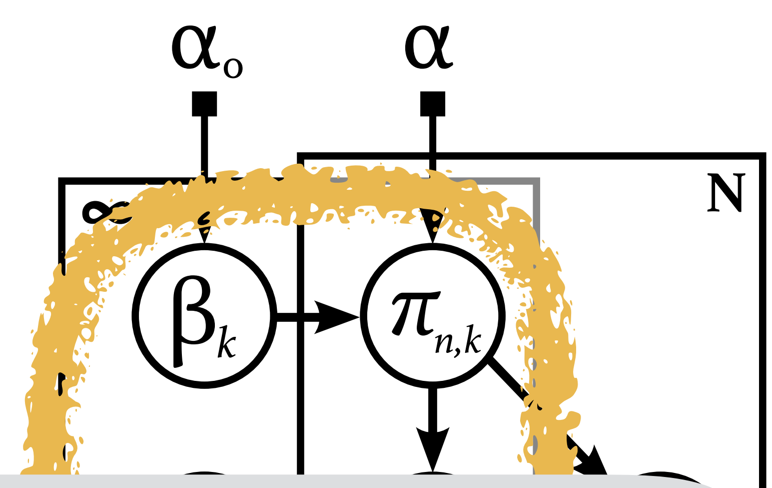
Bryant and Sudderth, 2012

K

consider splitting each factor

conside

some



ρ

$$\lambda[\pi_{n,k}]$$

$$\lambda^S[\pi_{n,k'}] = \rho_t \lambda[\pi_{n,k}] \qquad \lambda^S[\pi_{n,k''}] = (1-\rho_t)\lambda[\pi_{n,k}]$$

split/merge overview

Bryant and Sudderth, 2012

K

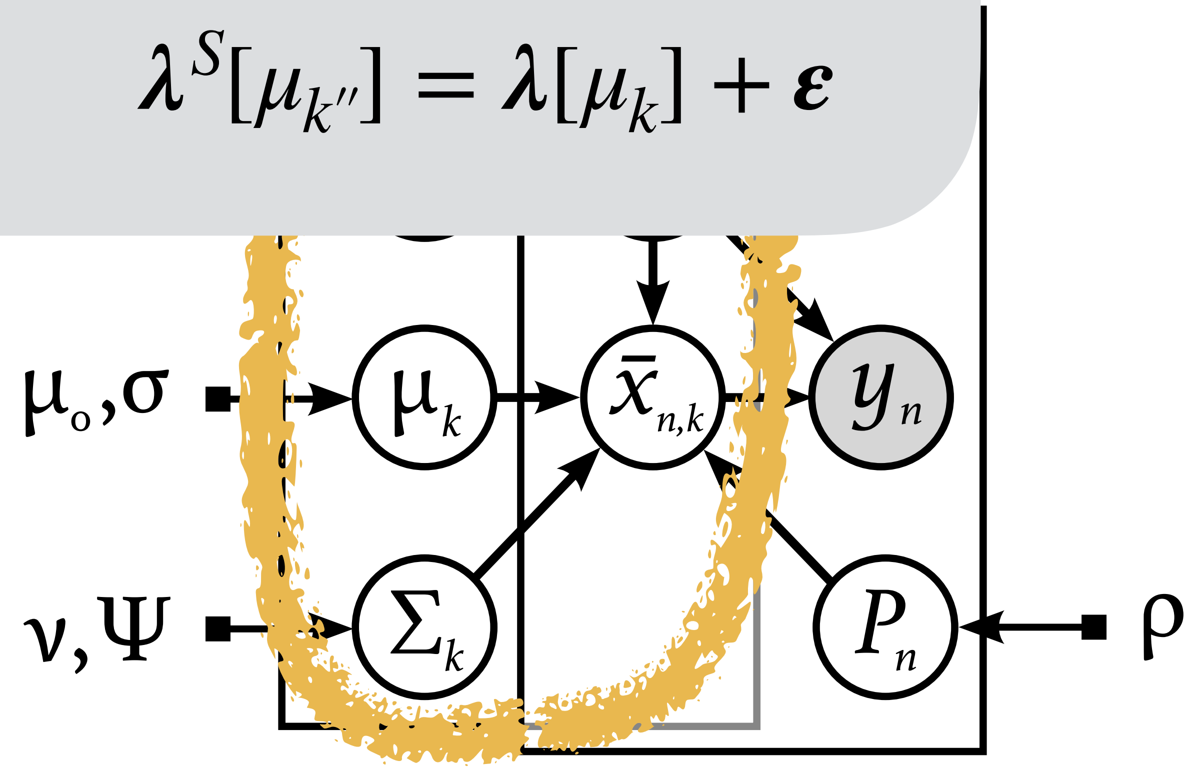
$$\lambda[\mu_k]$$

$$\lambda^S[\mu_{k'}] = \lambda[\mu_k] \qquad \lambda^S[\mu_{k''}] = \lambda[\mu_k] + \boldsymbol{\varepsilon}$$

consider

each

consider merging
some factors



split/merge overview

Bryant and Sudderth, 2012

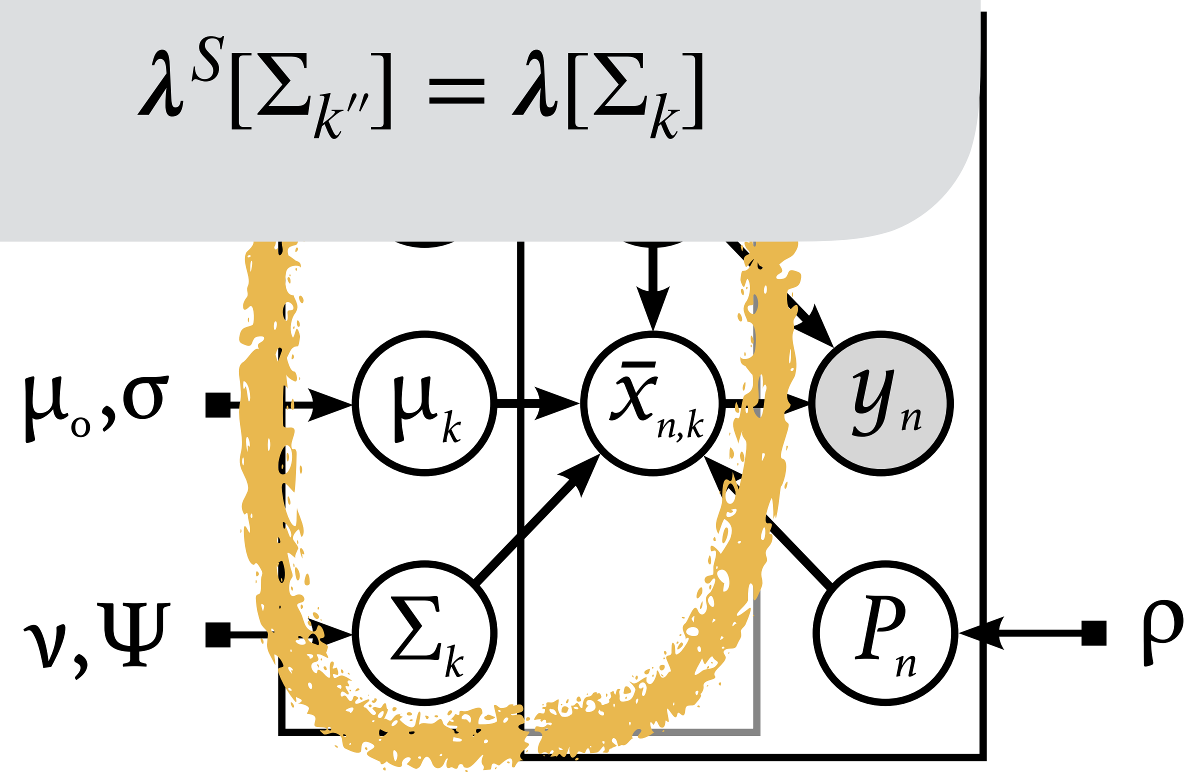
$$\lambda[\Sigma_k]$$

$$\lambda^S[\Sigma_{k'}] = \lambda[\Sigma_k] \qquad \lambda^S[\Sigma_{k''}] = \lambda[\Sigma_k]$$

K

consider
each

consider merging
some factors



split/merge overview

Bryant and Sudderth, 2012

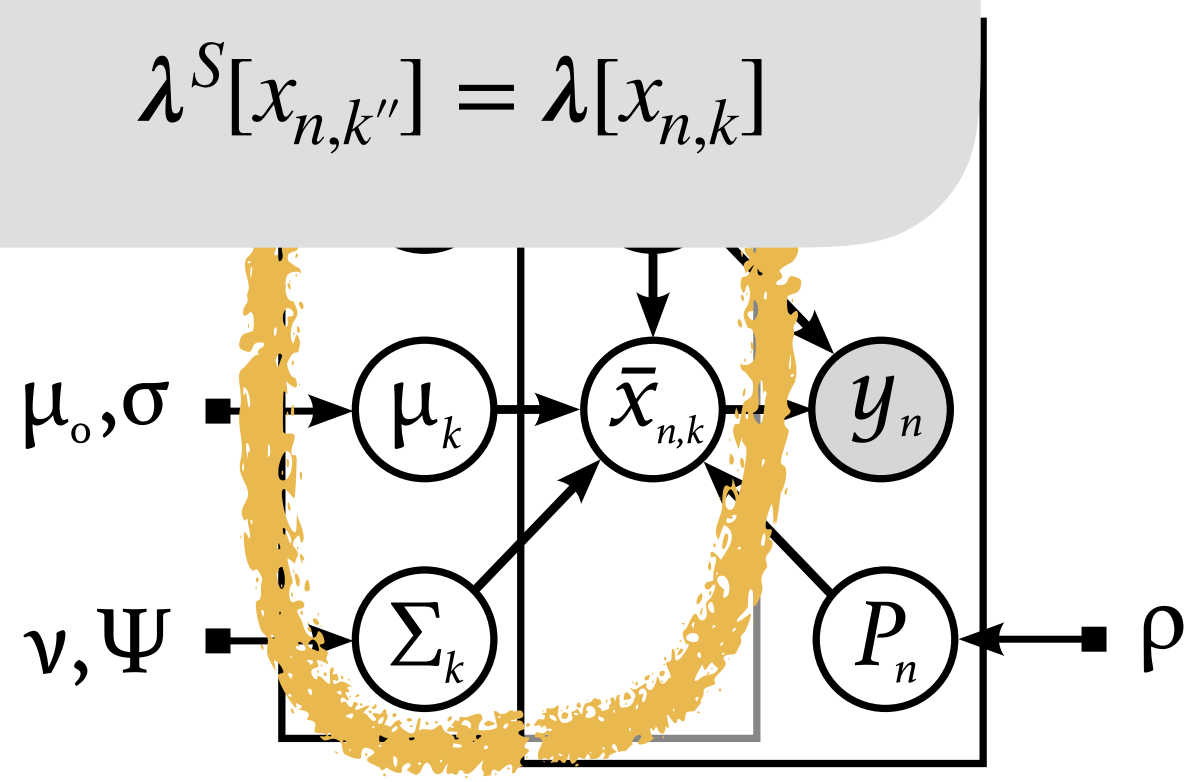
K

consider
each

$$\lambda[x_{n,k}]$$

$$\lambda^S[x_{n,k'}] = \lambda[x_{n,k}] \qquad \lambda^S[x_{n,k''}] = \lambda[x_{n,k}]$$

consider merging
some factors



split/merge overview

Bryant and Sudderth, 2012

K

consider

each

consider

some

$$\lambda^M[\beta_k] = \lambda[\beta_{k'}] + \lambda[\beta_{k''}]$$

$$\lambda^M[\pi_{n,k}] = \lambda[\pi_{n,k'}] + \lambda[\pi_{n,k''}]$$

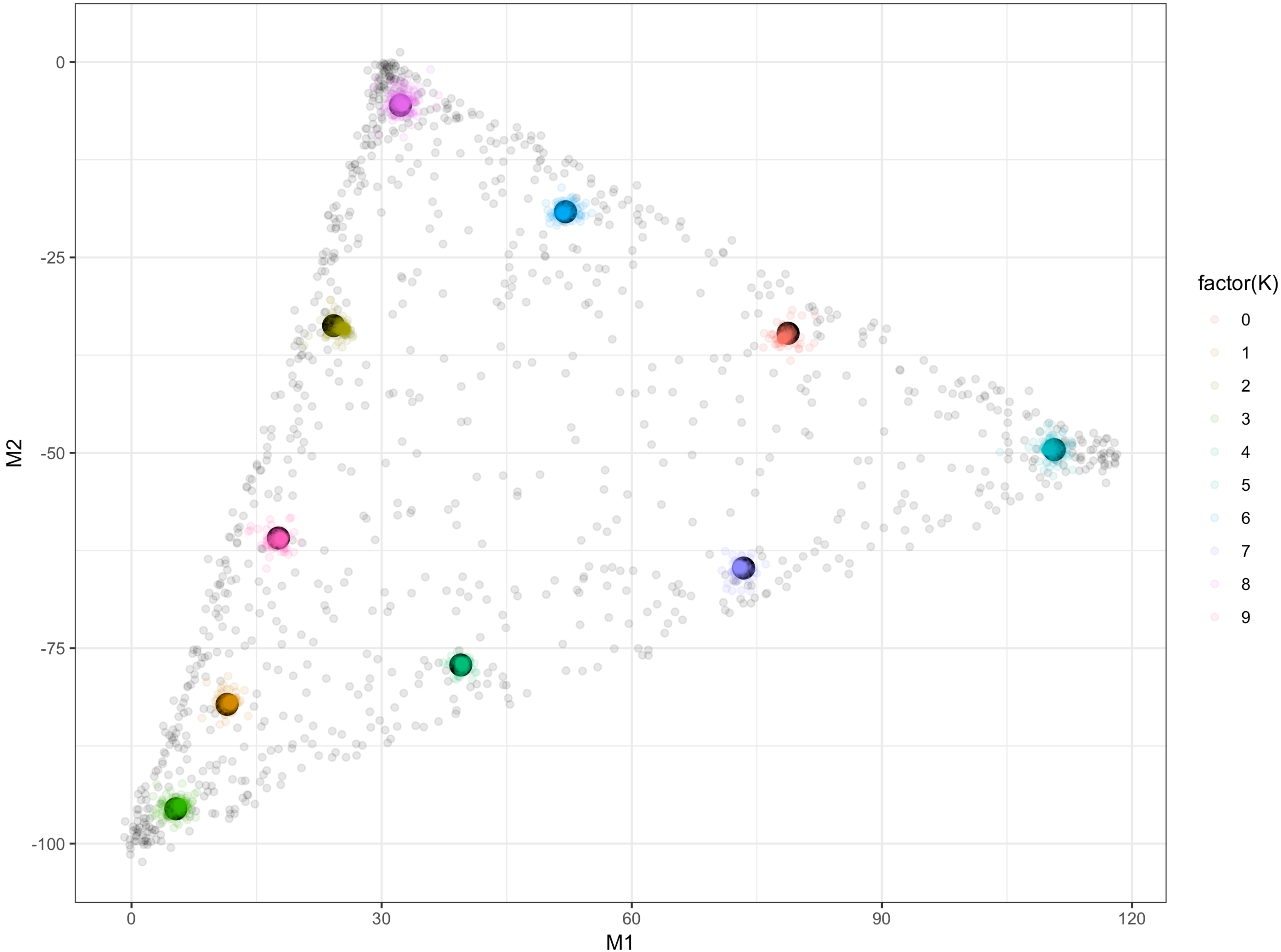
$$\lambda^M[\mu_k] = \frac{\lambda[\beta_{k'}]\lambda[\mu_{k'}] + \lambda[\beta_{k''}]\lambda[\mu_{k''}]}{\lambda[\beta_{k'}] + \lambda[\beta_{k''}]}$$

$$\ldots$$

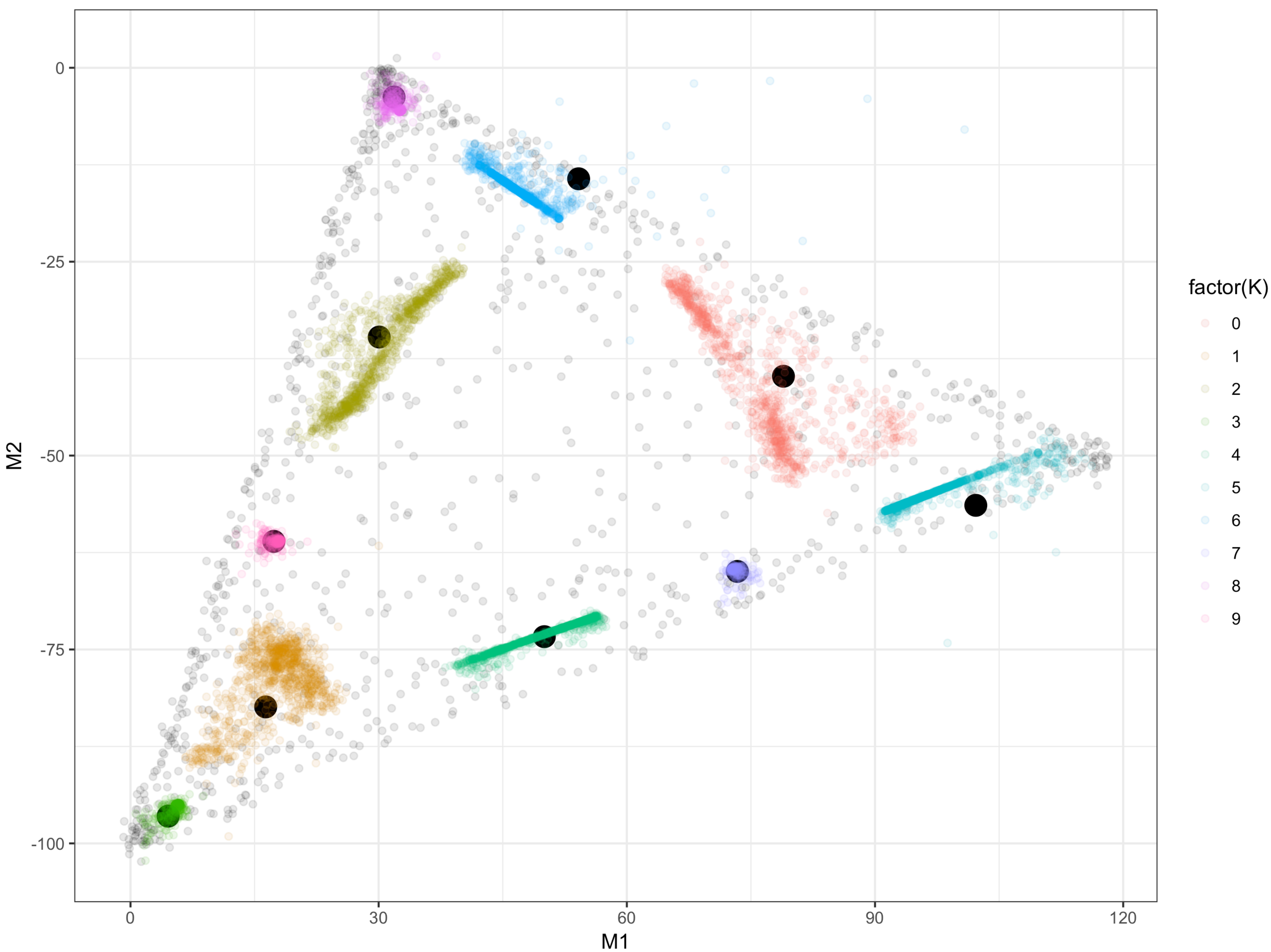
ρ

Algorithm Pseudocode

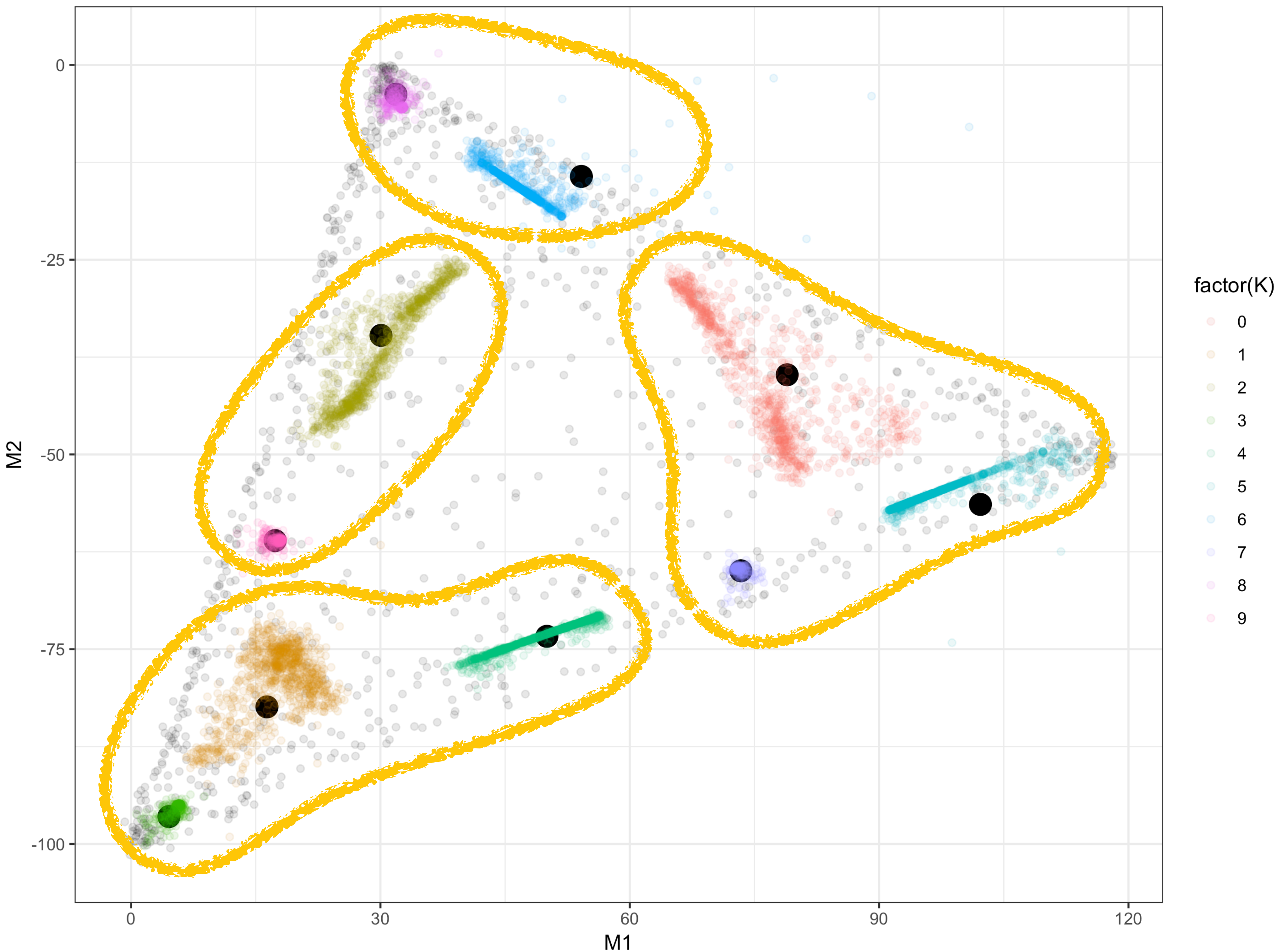
```
set K to an initial value
initialize variational parameters
repeat until convergence:
    repeat until batch convergence:
        update variational parameters for
          x̄, π, P, β using BBVI
        update variational parameters for
          μ, Σ using analytic updates
    split/merge latent factors, defining new K
      and updating variational parameters
      accordingly
```



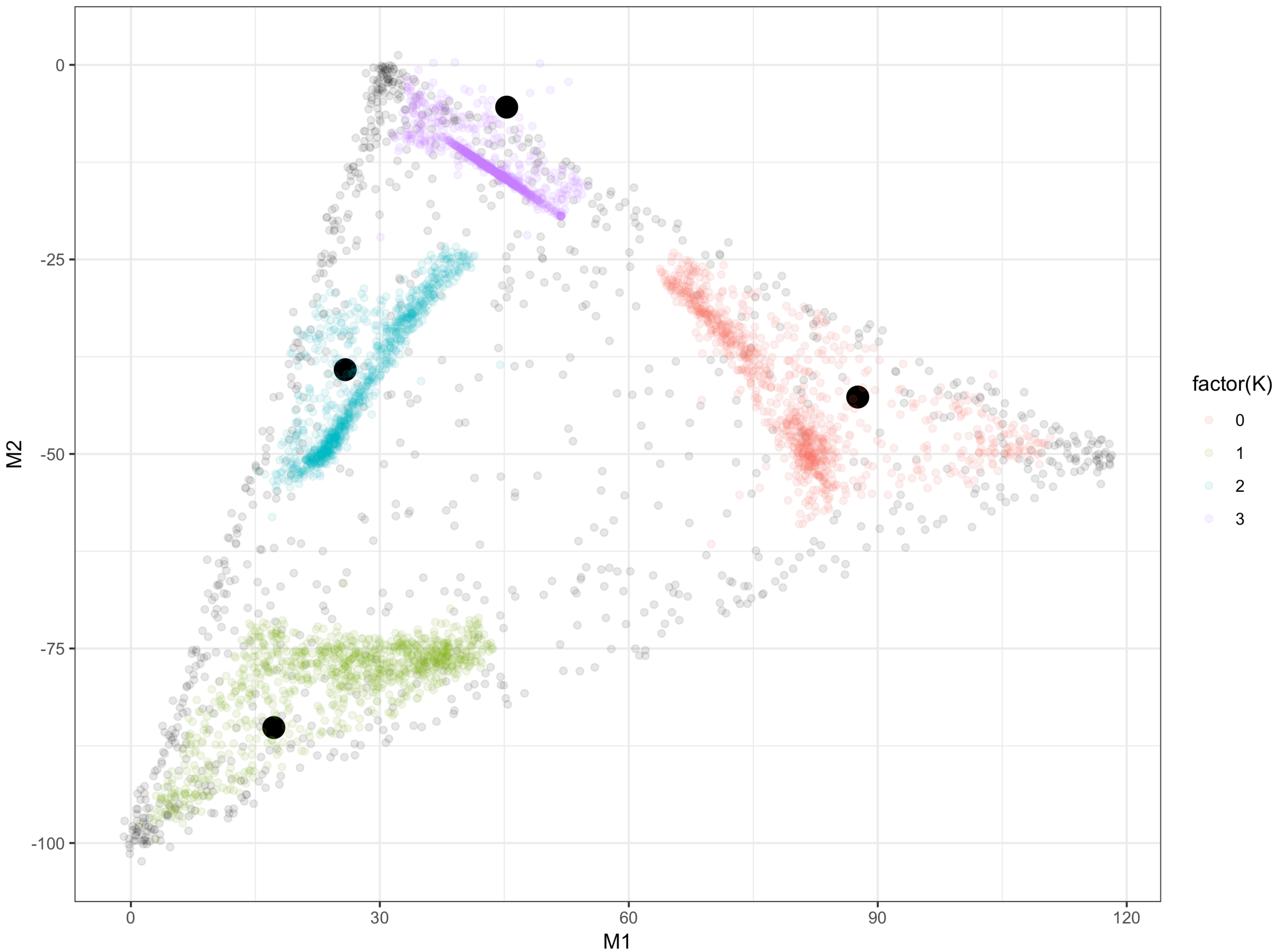
M2
0
-25
-50
-75
-100
M1
0
30
60
90
120
factor(K)
0
1
2
3
4
5
6
7
8
9



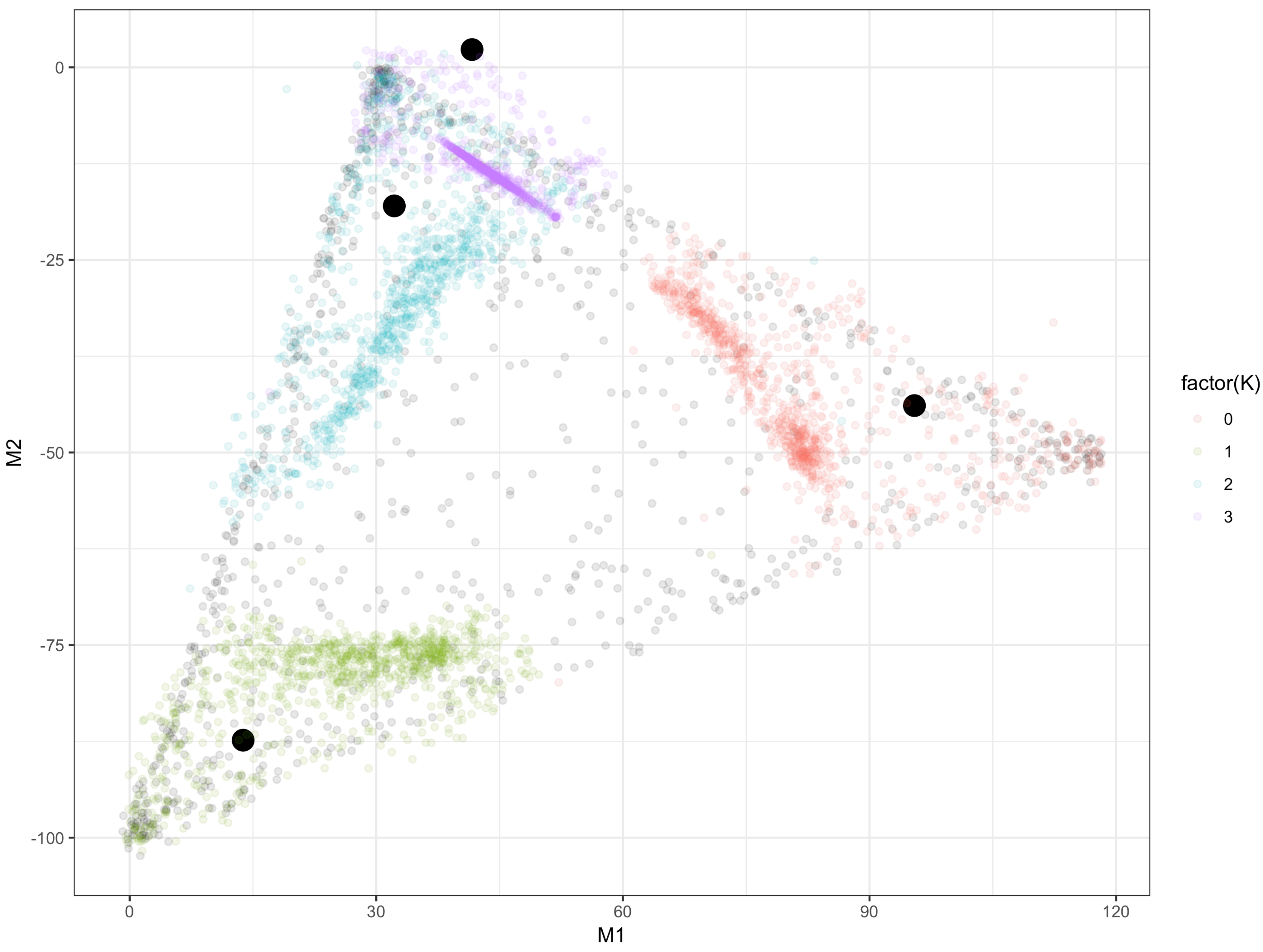
0
-25
-50
-75
-100
M2
0
30
60
90
120
M1
factor(K)
0
1
2
3
4
5
6
7
8
9



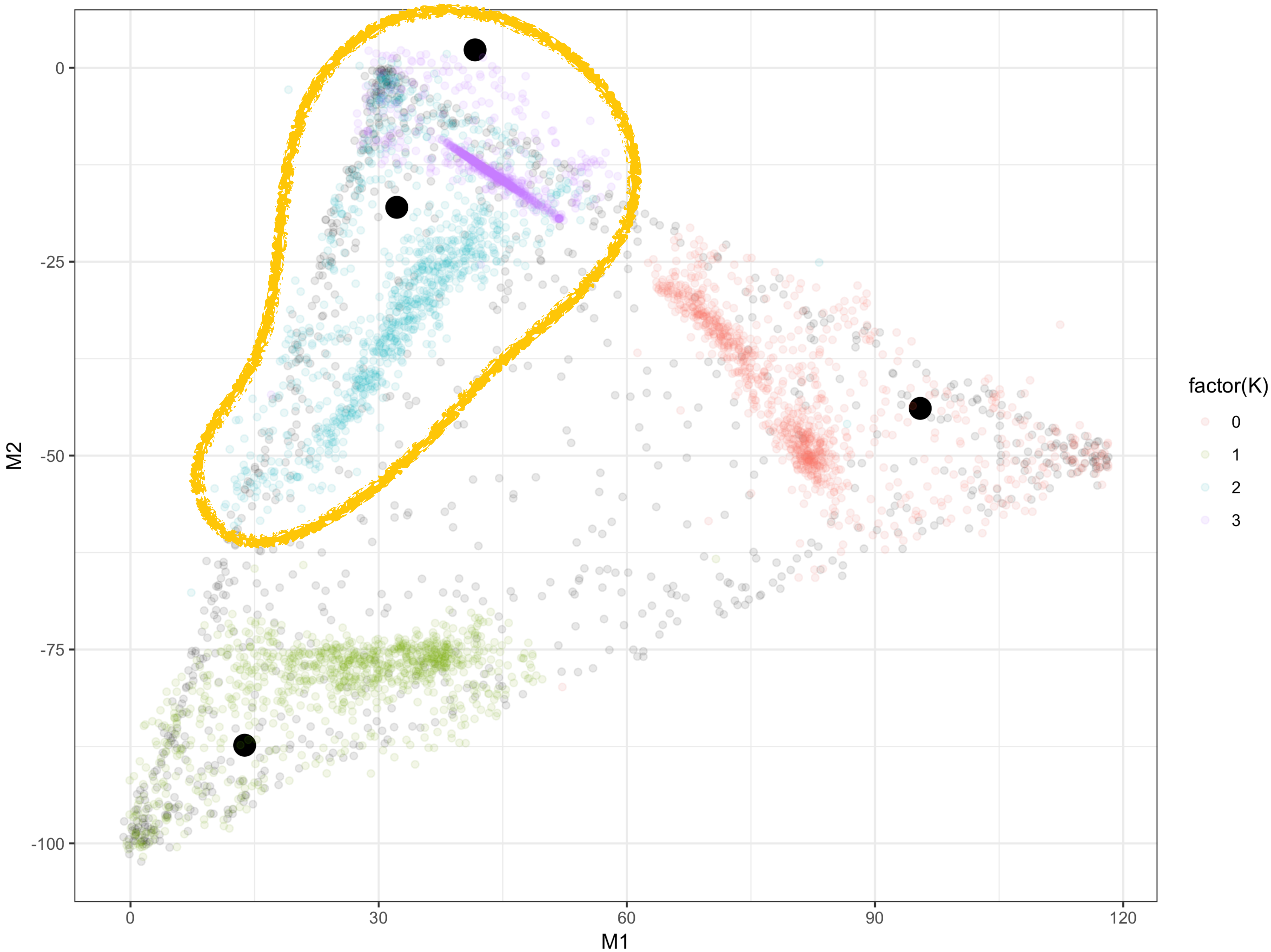
0
-25
-50
-75
-100
0
30
60
90
120
M1
M2
factor(K)
0
1
2
3
4
5
6
7
8
9



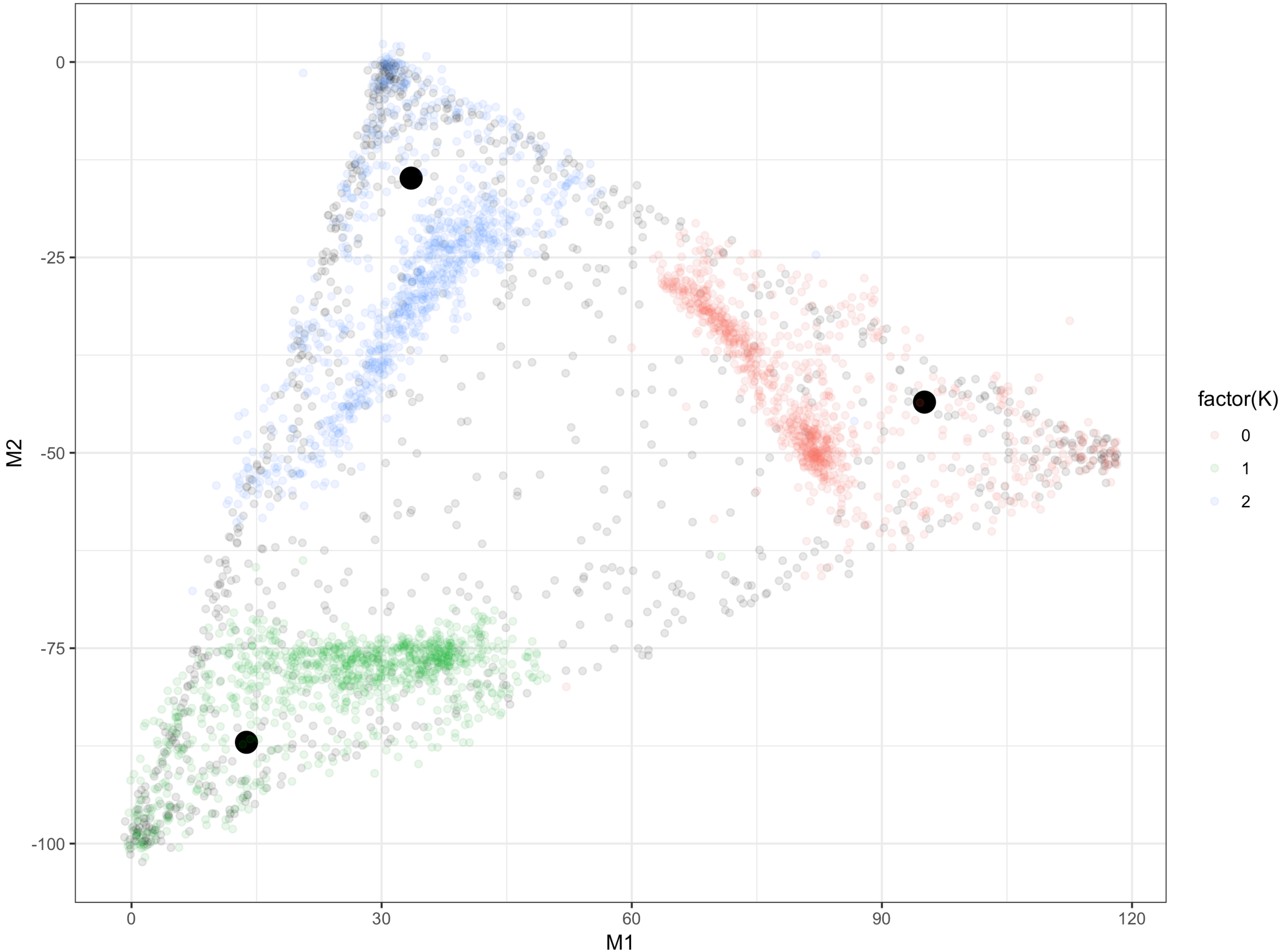
0
-25
-50
-75
-100
M2
0
30
60
90
120
M1
factor(K)
0
1
2
3



0
-25
-50
-75
-100
M2
0
30
60
90
120
M1
factor(K)
0
1
2
3

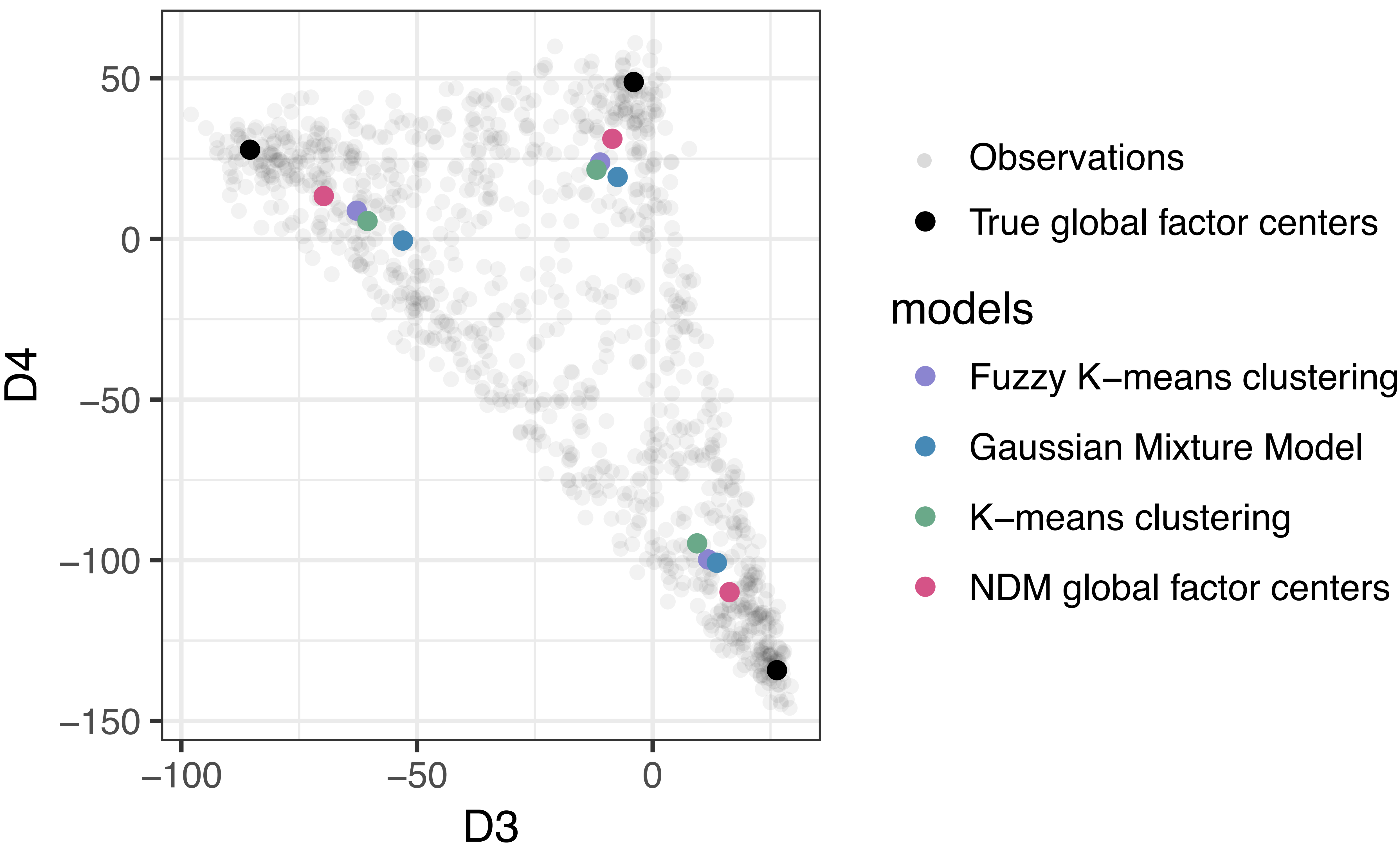


0
-25
-50
-75
-100
0
30
60
90
120
M1
M2
factor(K)
0
1
2
3

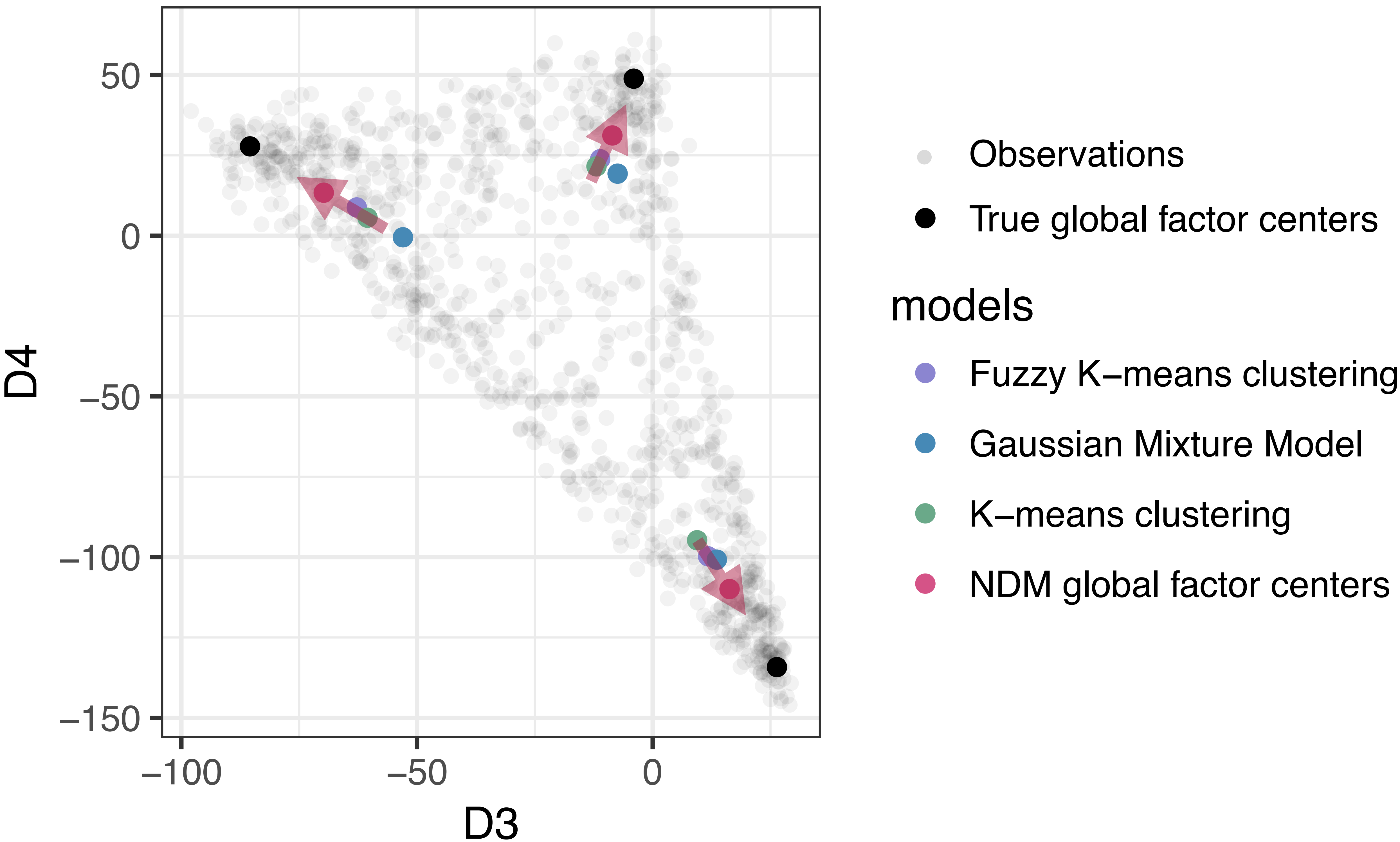


0
-25
-50
-75
-100
M2
0
30
60
90
120
M1
factor(K)
0
1
2

Results on Simulated Data

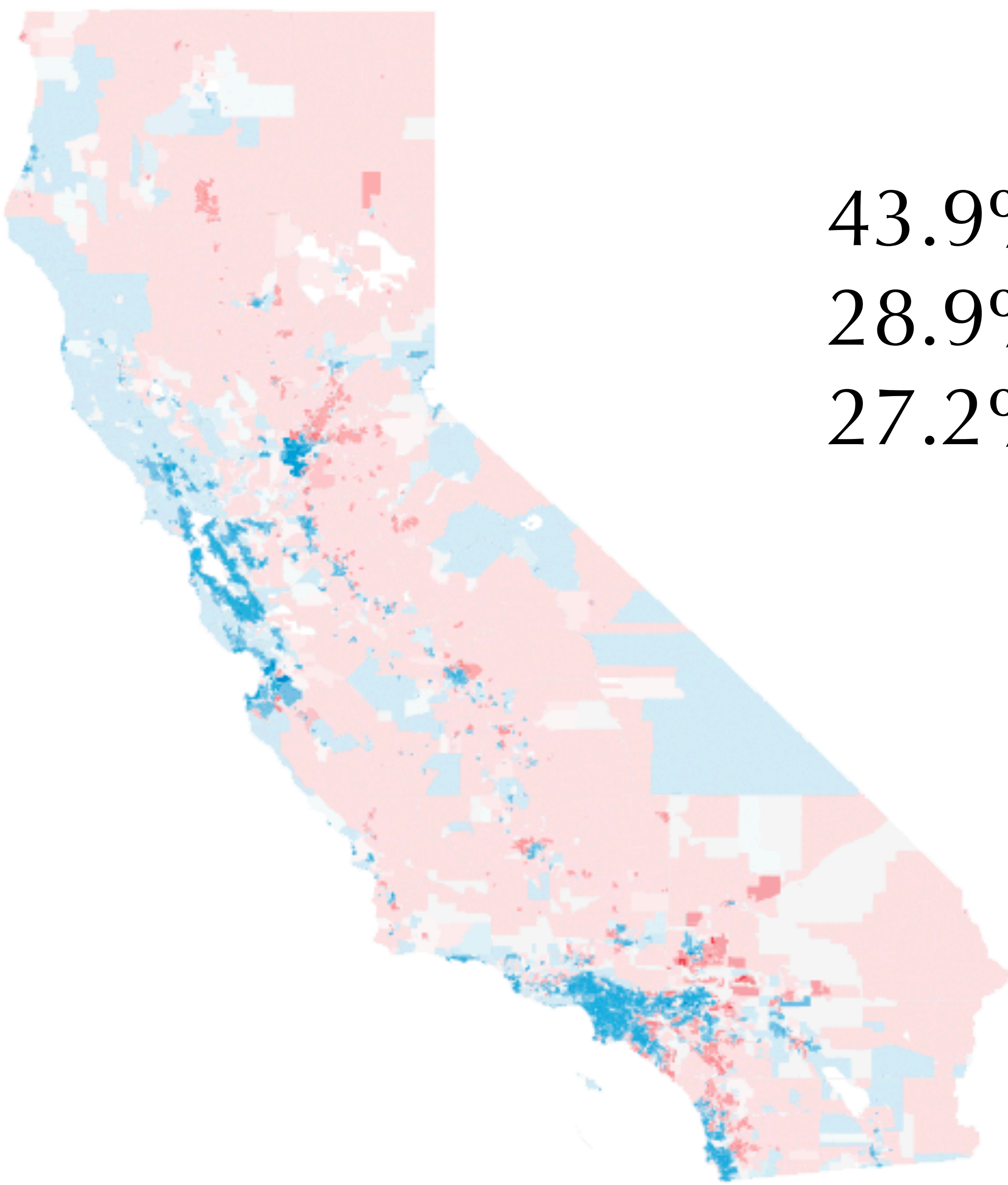


Results on Simulated Data



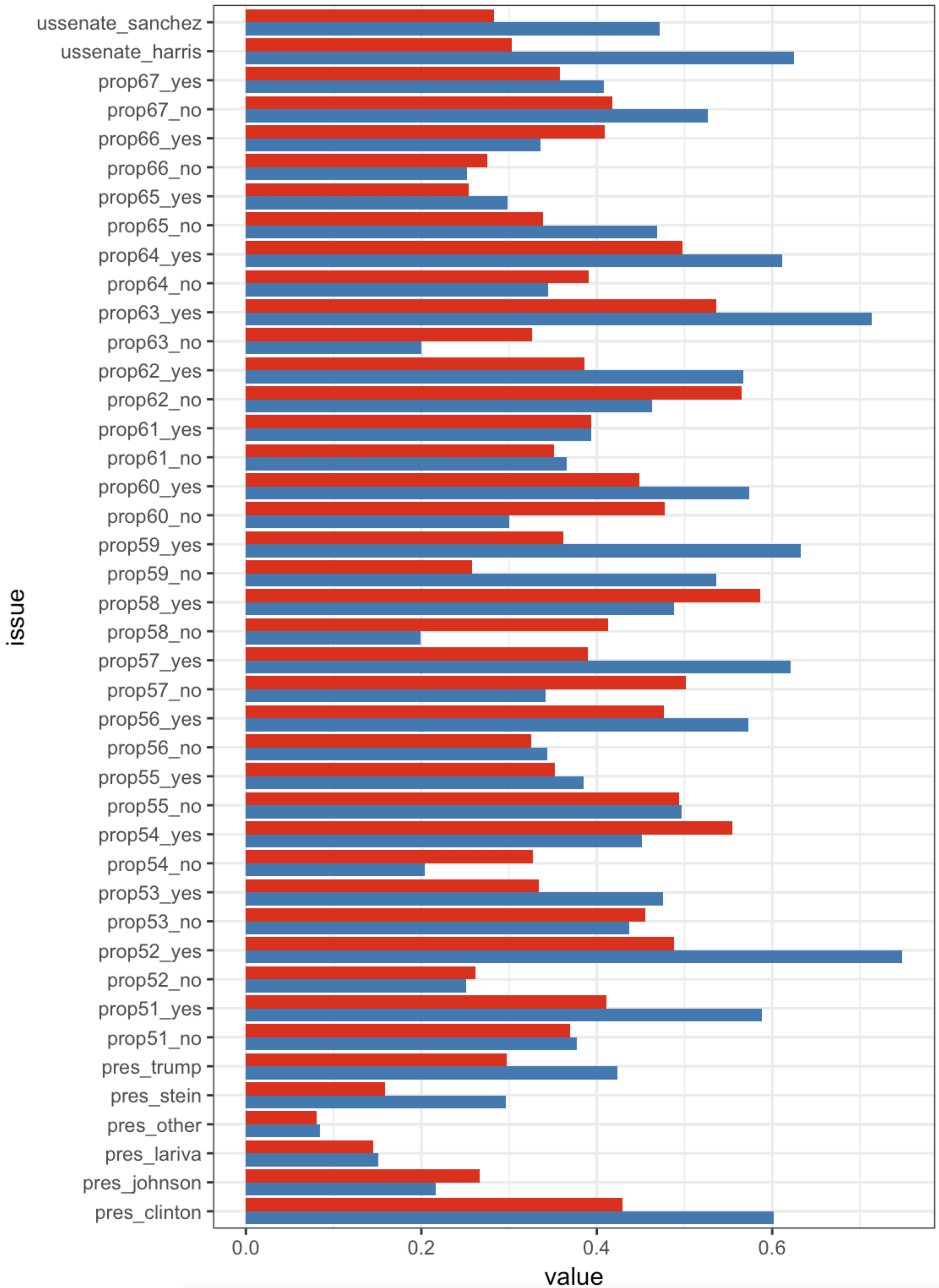
2016 Election in California

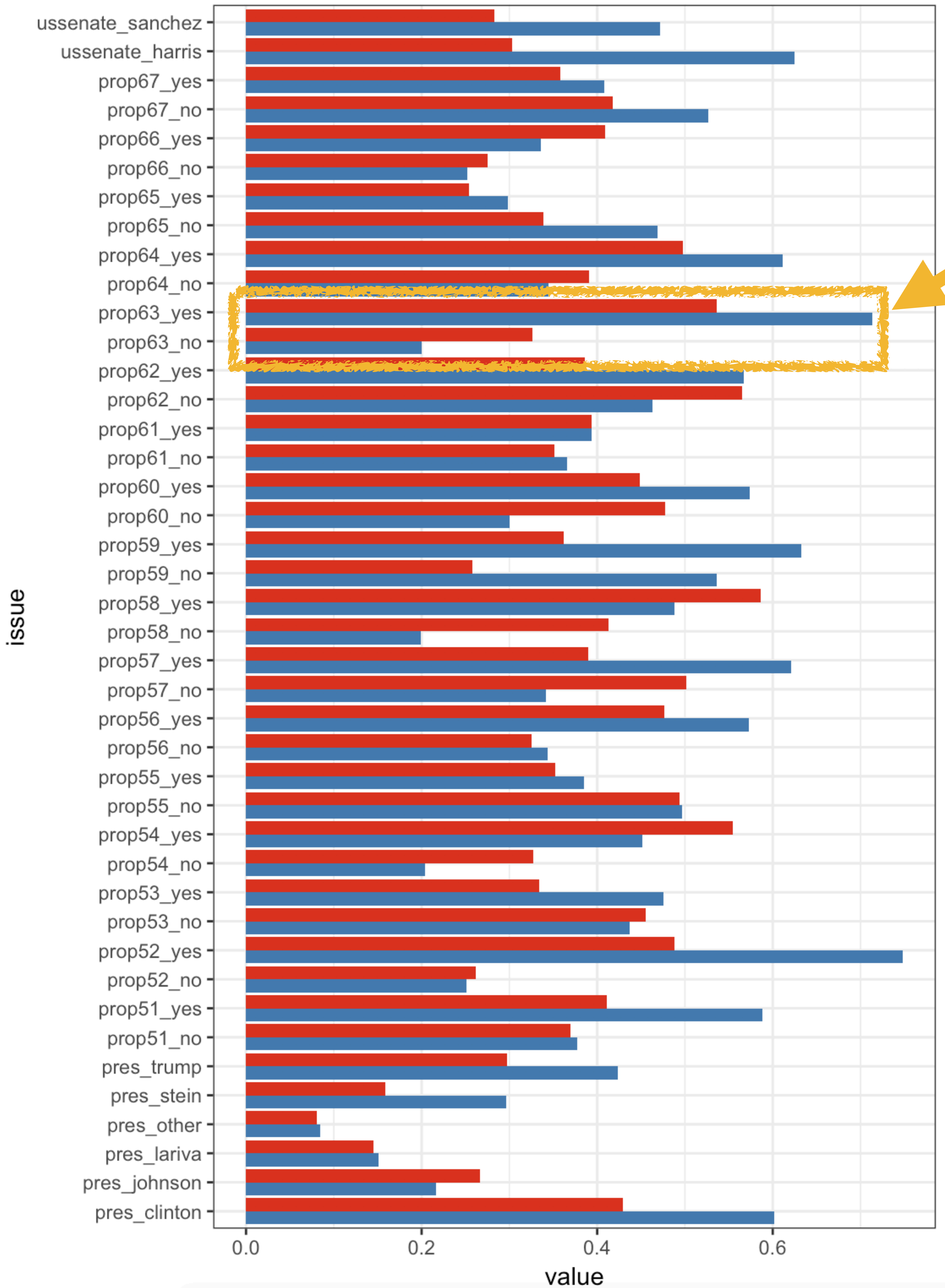
https://github.com/datadesk/california-2016-election-precinct-maps



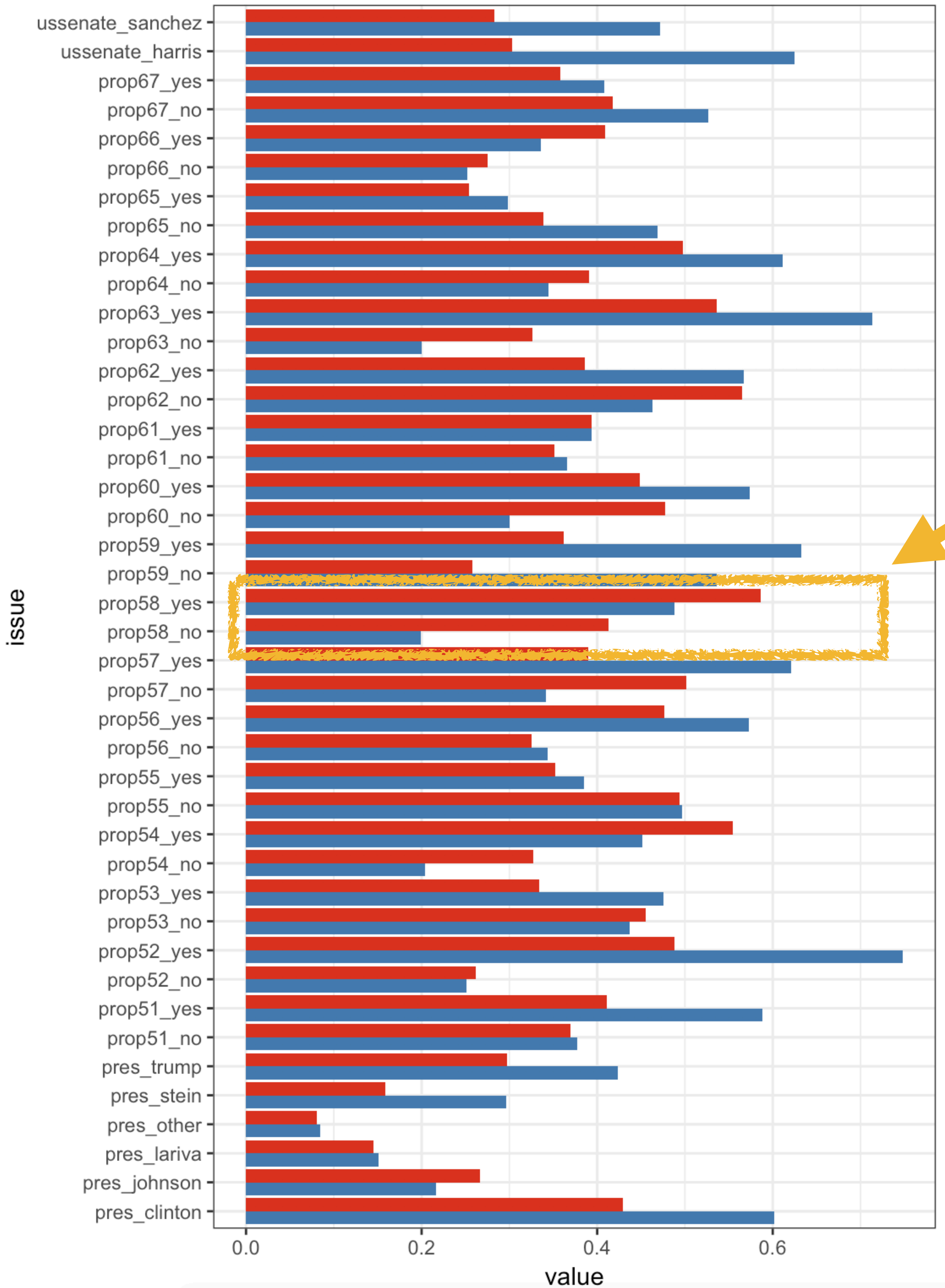
43.9% registered Democrats
28.9% registered Republicans
27.2% other parties / unregistered

caveat: these are
very preliminary results

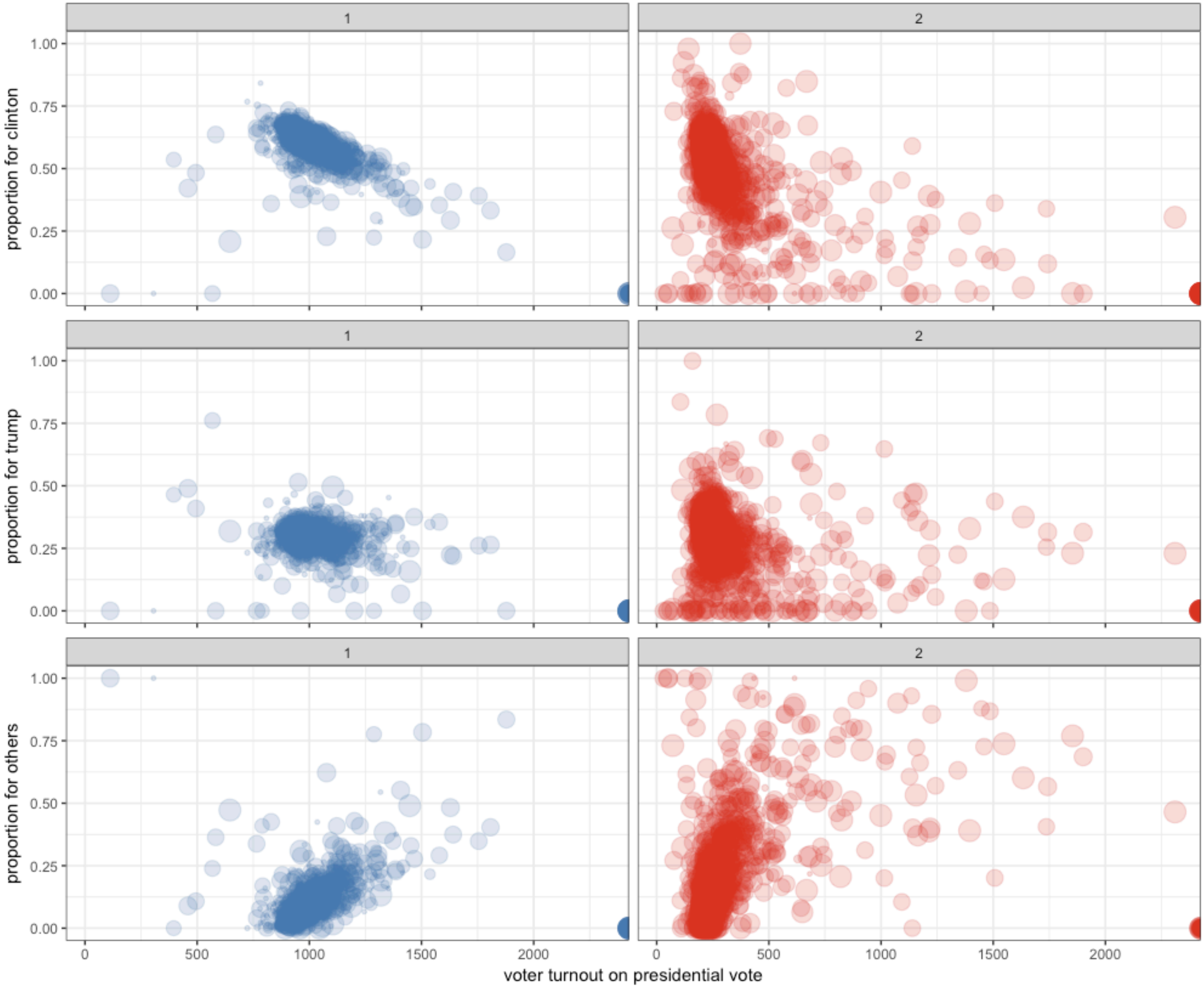

ussenate_sanchez
ussenate_harris
prop67_yes
prop67_no
prop66_yes
prop66_no
prop65_yes
prop65_no
prop64_yes
prop64_no
prop63_yes
prop63_no
prop62_yes
prop62_no
prop61_yes
prop61_no
prop60_yes
prop60_no
prop59_yes
prop59_no
prop58_yes
prop58_no
prop57_yes
prop57_no
prop56_yes
prop56_no
prop55_yes
prop55_no
prop54_yes
prop54_no
prop53_yes
prop53_no
prop52_yes
prop52_no
prop51_yes
prop51_no
pres_trump
pres_stein
pres_other
pres_lariva
pres_johnson
pres_clinton
issue
0.0
0.2
0.4
0.6
value



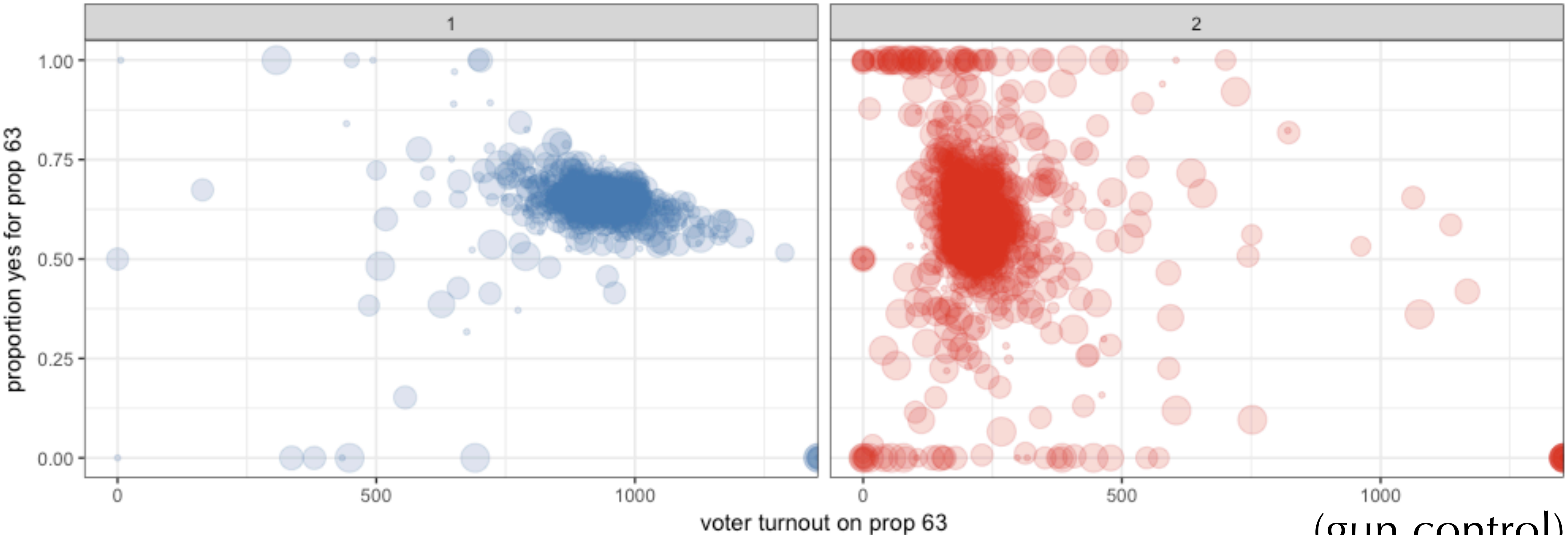
Prop 63: Background Checks for Ammunition Purchases and Large-Capacity Ammunition Magazine Ban



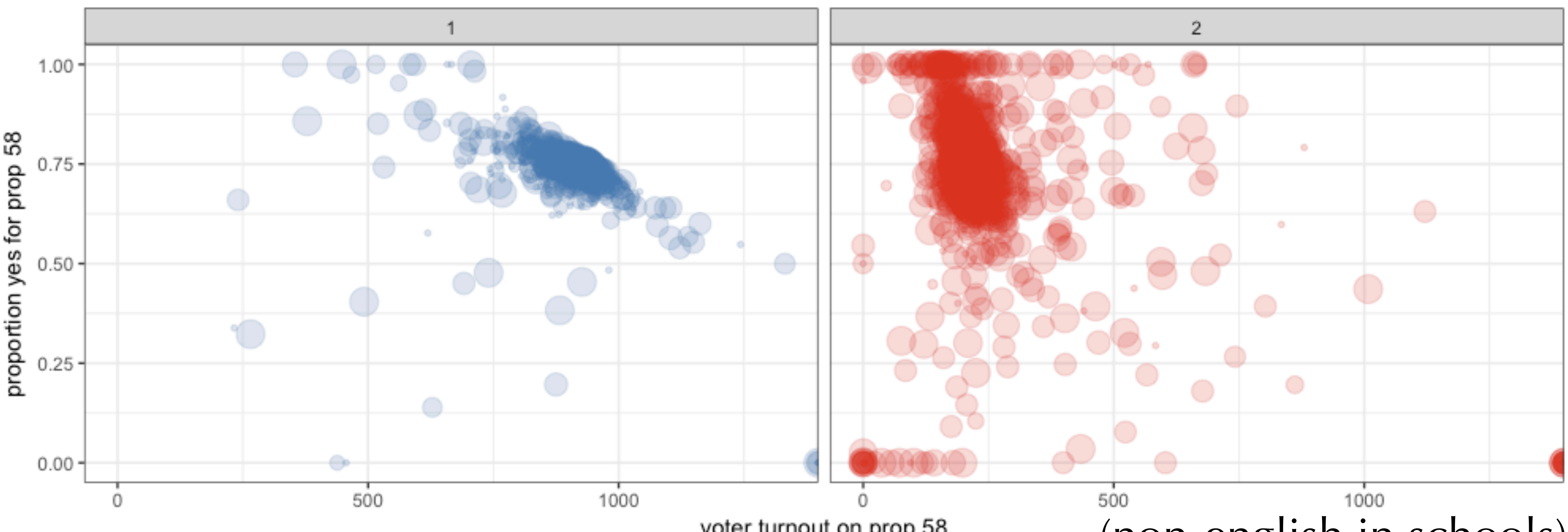
Prop 58: Non-English Languages Allowed in Public Education



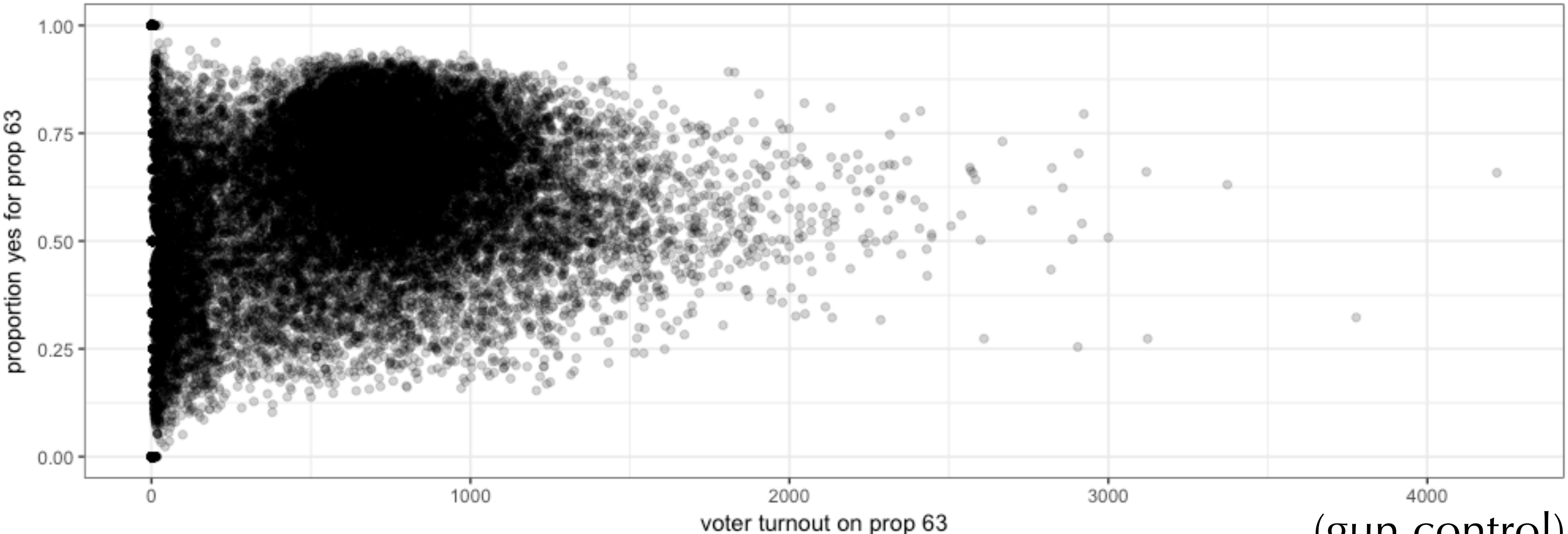
1
2
proportion for clinton
proportion for trump
proportion for others
1.00
0.75
0.50
0.25
0.00
0
500
1000
1500
2000
voter turnout on presidential vote



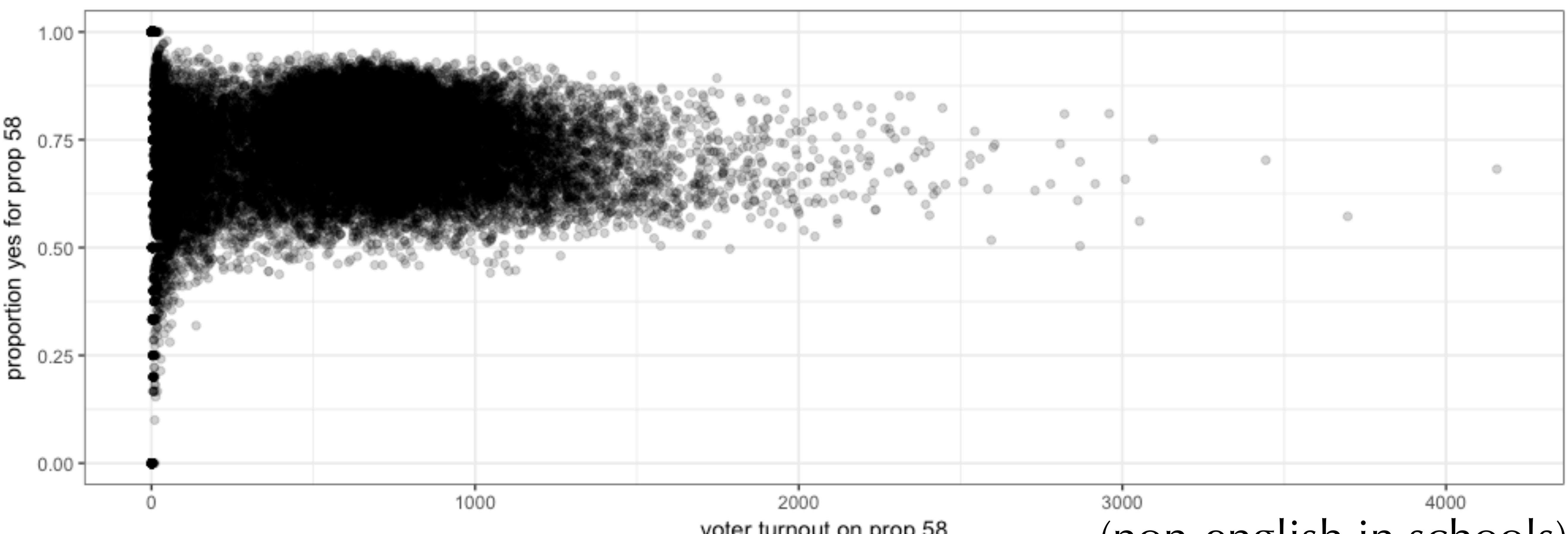
(gun control)



(non-english in schools)



(gun control)



(non-english in schools)

Thank you!

This research was supported by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at Princeton University, administered by Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence.