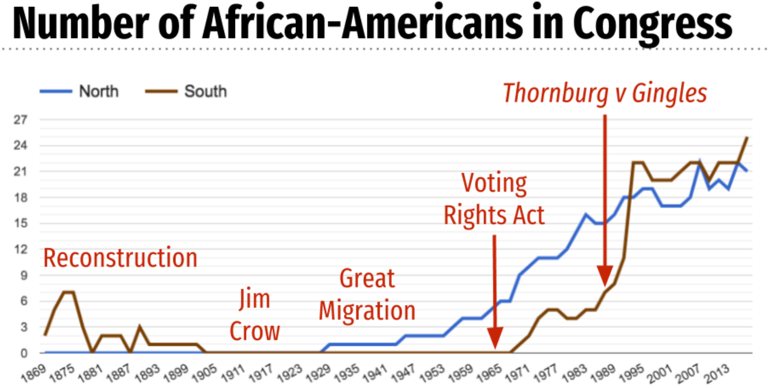
Flexible Ecological Inference

Jameson Quinn

3/1/2019

Teaser



Number of African-Americans in Congress
North
South
Thornburg v Gingles
Voting
Rights Act
Reconstruction
Jim
Crow
Great
Migration
27
24
21
18
15
12
9
6
3
0
1869 1875 1881 1887 1893 1899 1905 1911 1917 1923 1929 1935 1941 1947 1953 1959 1965 1971 1977 1983 1989 1995 2001 2007 2013

Thornburg v Gingles, 1986

A majority-minority district must be created if:

1. A minority group is "sufficiently numerous and compact to form a majority in a single-member district"; and
2. The minority group is **"politically cohesive"**; and
3. The "majority **votes sufficiently as a bloc** to enable it . . . usually to defeat the minority's preferred candidate."

Ecological data

		Candidate			
		R	*D*	*No*	
Race	*White*	?	?	?	400
	Black	?	?	?	200
	Hispanic	?	?	?	100
	Other	?	?	?	100
		400	200	200	800

precinct

Majority=Majority?

Race		Candidate			
		R	*D*	*No*	
	White	400	0	0	400
	Black	0	200	0	200
	Hispanic	0	0	100	100
	Other	0	0	100	100
		400	200	200	800

precinct

Backwards?

Race		Candidate			
		R	*D*	*No*	
	White	0	200	200	400
	Black	200	0	0	200
	Hispanic	100	0	0	100
	Other	100	0	0	100
		400	200	200	800

precinct

Independence?

	Candidate			
Race	*R*	*D*	*No*	
White	200	100	100	400
Black	100	50	50	200
Hispanic	50	25	25	100
Other	50	25	25	100
	400	200	200	800

precinct

Flexible model (1)

$$\vec{y}_{p,r} = y_{p,r,c}\|_{c=1}^{C} \sim \text{Multinomial}\left(n_{p,r}, \frac{exp(\cdots)\|_{c=1}^{C}}{\sum_{c=1}^{C} exp(\cdots)}\right)$$

Want differentiability, for observed information. So, replace Multinomial with CMult. This adds a small amount of bias; discussion of this issue beyond the scope of this talk.

Flexible model (2)

$$\vec{y}_{p,r} = y_{p,r,c}\|_{c=1}^{C} \sim \text{Cmult}\left(n_{p,r}, \frac{exp(\alpha_c + \beta_{r,c})\|_{c=1}^{C}}{\sum_{c=1}^{C} exp(\alpha_c + \beta_{r,c})}\right)$$

$$\alpha_c \sim \mathcal{N}(0, \sigma_\alpha) \qquad \sigma_\alpha \sim \text{Expo}(5)$$

$$\beta_{r,c} \sim \mathcal{N}(0, \sigma_\beta) \qquad \sigma_\beta \sim \text{Expo}(5)$$

xmastree_bare.jpg

Flexible model (3)

$$\vec{y}_{p,r} = y_{p,r,c}\|_{c=1}^{C} \sim \text{CMult}\left(n_{p,r}, \frac{exp(\alpha_c + \beta_{r,c} + \lambda_{r,c,p})\|_{c=1}^{C}}{\sum_{c=1}^{C} exp(\alpha_c + \beta_{r,c} + \lambda_{r,c,p})}\right)$$

$$\alpha_c \sim \mathcal{N}(0, \sigma_\alpha) \qquad \sigma_\alpha \sim \text{Expo}(5)$$

$$\beta_{r,c} \sim \mathcal{N}(0, \sigma_\beta) \qquad \sigma_\beta \sim \text{Expo}(5)$$

$$\lambda_{r,c,p} \sim \mathcal{N}(0, \sigma_\beta) \qquad \sigma_\lambda \sim \text{Expo}(5)$$

λ handles overdispersion. Note: Bayesian Occam's Razor.

Standard Bayesian approach (simplified)

$$\begin{array}{c}\text{priors } \pi \\ \downarrow \\ \text{parameters } \theta \\ \downarrow \\ \text{latentvariables } y \\ \downarrow \\ \text{data } z\end{array}$$

Standard Bayesian approach (cont'd)

priors π	
$\downarrow$	(parameterizes distribution)
parameters θ	(unobservable quantities of interest; low-D)
$\downarrow$	(parameterizes distribution)
latent variables y	(unobservable nuisance parameters; high-D)
$\downarrow$	(parameterizes distribution)
data z	

Ecological Inference

priors π	
$\downarrow$	(parameterizes distribution)
parameters θ	(unobservable nuisance parameters; low-D?)
$\downarrow$	(parameterizes distribution)
latent variables y	(unobserved quantities of interest; high-D)
$\downarrow$	(deterministic function)
data z	

A likelihood from a deterministic function is an indicator function!

Variational inference

Goal: approximate unnormalized posterior density $p(\theta, y|z) \propto p(z|\theta, y)p_\pi(\theta, y)$ with sampleable parametric distribution $q_\phi(\theta, y)$.

Maximize negative K-L divergence from approximation to normalized posterior $p(z|\theta, y)p_\pi(\theta, y)/p(z)$:

$$E_{q_\phi}\left(\log \frac{p(z|\theta, y)p_\pi(\theta, y)}{q_\phi(\theta, y)p(z)}\right) < 0$$

$$E_{q_\phi}\left(\log[p(z|y)p(\theta, y)] - \log[q_\phi(\theta, y)] - \log(p(z))\right) < 0$$

$$E_{q_\phi}\left(\log[p(z|y)p(\theta, y)] - \log[q_\phi(\theta, y)]\right) < \log(p(z))$$

LHS is "ELBO"; goal is to find ϕ which maximizes it.

ELBO terms

$E_{q_\phi}(\log[p(z|y)p(\theta, y)])$ is "energy" term. Maximized if q is a δ (dirac mass) at MLE for $(\theta, y|z)$. Unboundedly negative if q has probability mass where p doesn't.

$E_{q_\phi}(-\log[q_\phi(\theta, y)])$ is "entropy" term. Maximized by making q "diffuse"; for example, if q is $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, then this is inversely proportional to $\det(\Sigma)$. In principle unboundedly negative; in practice, easier to control than "energy" term.

Together, they're maximized if q_ϕ "imitates" p.

Ecological case

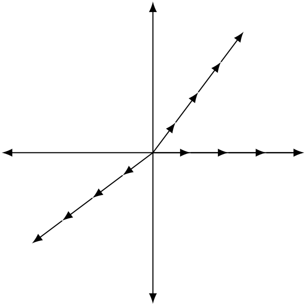
Define $\mathcal{Y}_z : \{y \in \mathbb{R}^{RC} : p(z|y) = 1\}$.

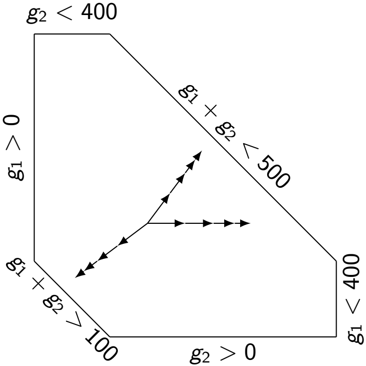
Use observed data z to construct (diffeomorphic) $g(y') : \mathbb{R}^{(R-1)(C-1)} \to \mathcal{Y}_z$. Choose $q_{\phi,z}(\theta, y')$ with full support over $y' \in \mathbb{R}^{(R-1)(C-1)}$.

Now, $p[z|g(y')] = 1$, so ELBO simplifies to:

$$E_{q_\phi}\left(\log[p(\theta, g(y'))] - \log[q_\phi(\theta, y')]\right)$$

How to construct $g(y')$?





$$\frac{\partial \sqrt{g(y')^T g(y')}}{\partial \sqrt{y'^T y'}} = \frac{1}{g_1 g_2 (400 - g_1)(400 - g_2)(500 - g_1 - g_2) \cdots}$$

Choose the form of your posterior

observed information

Lower-D posterior

amortization

Thanks

Thanks