Introduction to Ecological Inference

Yuki Atsusaka

Department of Political Science
Rice University

This material was prepared for POLI 592 (Advanced Political Methodology)

First version: April 11, 2019
Current version: April 15, 2019

What is Ecological Inference?

Ecological inference is the process of

- Learning about **individual** level behavior from **aggregate** information
- Learning about unknown **joint** distribution from the combination of known **marginal** distributions

Motivation

- Voting Rights Act Litigation
  - Courts want to know if people vote differently by race (Greiner 2009; Greiner 2011)
- Disease Prevention and Management
  - Epidemiologists want to know which age group has the highest HIV prevalence (Wakefield 2004)
- Political Behavior
  - Party students want to know how much split-ticket voting occurs (Johnston & Pattie 2003; Benoit, Laver, & Giannetti 2004; Park, Hanmer, & Biggers 2014; Plescia & De Sio 2018)
  - Race scholars want to know if minority turnout increases as more minority descriptive representation grows (Gay 2001; Heron and Sekhon 2005; Hajnal & Troustine 2005; Baretto 2007)

Motivation

- Historical work
  - Comparativists want to evaluate the policy to boost female turnout in 1920s Germany (Homola n.d.)
  - Historians want to know which economic group supported the Nazi Party (King, Rosen, Tanner, & Wagner 2008; O'Loughlin 2000)
- Other contexts
  - Conflict scholars want to find which subpopulation join extremist groups in the world
  - IR researchers want to infer whether all countries in each bloc agreed on int'l climate agreements (Sprinz 2000)

For these research, often no individual level data is available or social desirable bias prevails

$\Rightarrow$ **Aggregate data as alternative options**

Ecological Fallacy

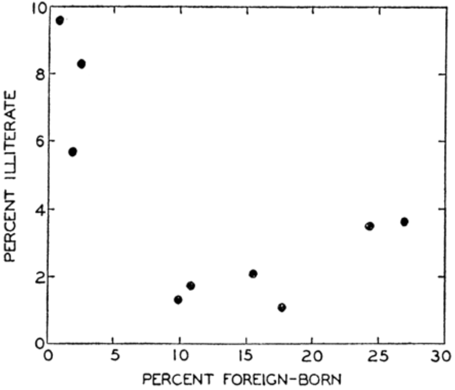
What can we say from aggregate data?

- Descriptive Observation
  $\rightsquigarrow$ At the county level, the proportion of white voters is positively correlated with the proportion of Republican voters
- Naïve Inference
  $\rightsquigarrow$ White voters are more likely to vote for the Republican Party
- Ecological Fallacies (Selvin 1958)
  - Confusion between ecological correlations and individual correlations
  - Confusion between group average and total average
  - Simpson's paradox
  - Confusion between higher average and higher likelihood
- Aggregation Bias

Robinson's Paradox

Robinson (2009)[1950] "Ecological Correlations and the Behavior of Individuals"

Figure 3.

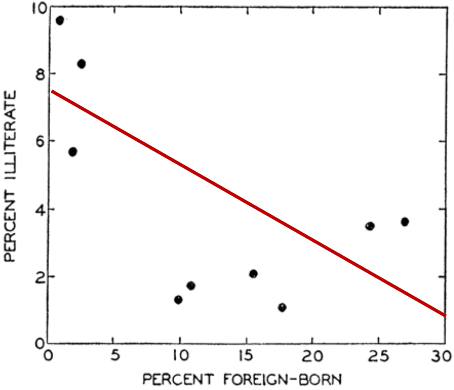


International Journal of Epidemiology, Volume 38, Issue 2, April 2009, Pages 337–341, https://doi.org/10.1093/ije/dyn357

Robinson's Paradox

Robinson (2009)[1950] "Ecological Correlations and the Behavior of Individuals"

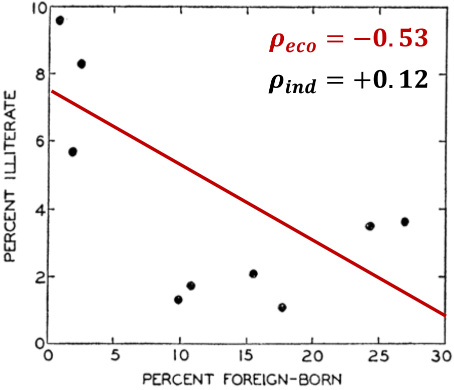
Figure 3.



International Journal of Epidemiology, Volume 38, Issue 2, April 2009, Pages 337–341, https://doi.org/10.1093/ije/dyn357

Robinson (2009)[1950] "Ecological Correlations and the Behavior of Individuals"

Figure 3.



International Journal of Epidemiology, Volume 38, Issue 2, April 2009, Pages 337–341, https://doi.org/10.1093/ije/dyn357

Aggregation Bias

Why Robinson's paradox? (Intuitively)

	Literate	Illiterate	
Foreign-born	110	50	**160**
US-born	30	30	**60**
	140	**80**	

Naïvely Inferred Pattern $\rightsquigarrow$ Foreign-born are more literate

	Literate	Illiterate	
Foreign-born	80	80	**160**
US-born	60	0	**60**
	140	**80**	

Another Possible Reality $\rightsquigarrow$ Foreign-born tend to live with literate people

$*$ Loss of information = indeterminacy problem

Aggregation Bias

Why Robinson's paradox? (Formally)

$$\text{cov}\left(\sum_{i=1}^{N} Y_i, \sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} \text{cov}(Y_i, X_i) + \underbrace{\sum_{i=1}^{N} \sum_{l \neq i} \text{cov}(Y_l, X_i)}_{\text{aggregation bias}} \tag{1}$$

⇝ When foreign-born ($X_i = 1$) is correlated with others' illiteracy ($Y_{l \neq i} = 1$), the second term is not zero (in this case negative)

⋆ In Goodman's regression, the correlation between the parameters (β_i^b, β_i^w) and X_i induces aggregation bias (King 1997, 40-41)

⇝ How can we do better than naïve inference?

What is Ecological Inference? (Review)

Ecological inference is the process of

- Learning about **individual** level behavior from **aggregate** information
- Learning about unknown **joint** distribution from the combination of known **marginal** distributions

Motivating Example

	Vote	Non Vote	
Black	N_{bV_i}	N_{bNV_i}	$\mathbf{N}_{b_i}$
White	N_{wV_i}	N_{wNV_i}	$\mathbf{N}_{w_i}$
	$\mathbf{N}_{V_i}$	$\mathbf{N}_{NV_i}$	

2×2 Contingency Table for Precinct i

- You have data from 100 precincts in Harris County
- You know # people who turned out from election returns **(Col marginals)**
- You know CVAP of whites and blacks from the US Census **(Row marginals)**
- You want to know how many whites/blacks went to vote (Internal cells)

Motivating Example (Extension)

	Clinton	Trump	Sanders	Non Vote	
White	N_{wB_i}	N_{wL_i}	N_{wK_i}	N_{wNV_i}	$\mathbf{N}_{w_i}$
Black	N_{bB_i}	N_{bL_i}	N_{bK_i}	N_{bNV_i}	$\mathbf{N}_{b_i}$
Asian	N_{aB_i}	N_{aL_i}	N_{aK_i}	N_{aNV_i}	$\mathbf{N}_{a_i}$
Hispanic	N_{hB_i}	N_{hL_i}	N_{hK_i}	N_{hNV_i}	$\mathbf{N}_{h_i}$
	$\mathbf{N}_{B_i}$	$\mathbf{N}_{L_i}$	$\mathbf{N}_{K_i}$	$\mathbf{N}_{NV_i}$	

R $\times$ C Contingency Table for Precinct i

History of EI Models

- Method of bounds or differences (Duncan & Davis 1953; Shively 1982; Achen & Shively 1995)
- Ecological regression (Ogburn & Goltra 1919; Goodman 1953, 1959; Freedman et al. 1991; Grofman & Barett 2009; Jiang, King, Schmaltz, & Tanner 2018)
- Combined approach (King 1997; Lewis 2004; Quinn 2004; Calvo & Escolar 2003)
- Bayesian hierarchical models (King, Rosen, & Tanner 1999; Rosen, Jiang, King, & Tanner 2001; Greiner and Quinn 2009; Wakefield 2004; Imai and Lu 2008)
- Machine learning variants (Flaxman 2015, 2016; Soldaini and YomTov 2016; Muezellec et al. 2017; Rosenman and Viswanathan 2018)

History of EI Models (Memorandum)

- Method of bounds or differences (Duncan & Davis 1953 [Method of bounds]; Shively 1982; Achen & Shively 1995 [Method of differences]))
- Ecological regression (Ogburn & Goltra 1919; Goodman 1953, 1959 [Ecological regression]; Freedman et al. 1991 [Neighborhood model]; Grofman & Barett 2009 [Double regression]; Jiang, King, Schmaltz, & Tanner 2018 [Partially identification]; ???[Homogeneous precinct analysis])
- Combined approach (King 1997 [King's EI]; Lewis 2004 [Seemingly unrelated EI]; Quinn 2004 [Dynamic EI]); Calvo & Escolar 2003 [Geographically weighted approach]
- Bayesian hierarchical models (King, Rosen, & Tanner 1999 [Binomial-Beta model]; Rosen, Jiang, King, & Tanner 2001 [Multinomial-Dirichlet model]; Greiner and Quinn 2009 [GQ model]; Wakefield 2004 [Wakefield's 2×2 model]; Imai and Lu 2008 [Multiple imputation model])
- Machine learning variants (Flaxman 2015, 2016 [Distributional regression model]; Soldaini and YomTov 2016 [Stochastic gradient descent model]; Muezellec et al. 2017 [Tsallis regularized optimal transport model]; Rosenman and Viswanathan 2018 [Poisson-Binomial model])

Binomial-Beta Model (King, Rosen, & Tanner 1999)

- What is Binomial-Beta model?
  - King's EI meets Markov chain Monte Carlo algorithms (Tanner 1996)
  - Covariate adjustment and its formal assessment
  - Robust to few observations and measurement error
  - "Borrowing strength" for efficiency and estimation
- Quantities of interest

	Vote	No Vote	
Black	β_i^b	$1-\beta_i^b$	X_i
White	β_i^w	$1-\beta_i^w$	$1-X_i$
	T_i	$1-T_i$	

  - Prop of blacks who voted: β_i^b $(= N_{BVi}/N_{Vi})$
  - Prop of whites who voted: β_i^w $(= N_{WVi}/N_{Vi})$
  - District-level props $B^b = \Sigma_{i=1}^{p} \frac{N_i^b \beta_i^b}{N^b}$; $B^w = \Sigma_{i=1}^{p} \frac{N_i^w \beta_i^w}{N^w}$

We estimate these unknown quantities from

- Known prop of blacks: X_i $(= N_{Bi}/N_i)$
- Known prop of turnout: T_i $(= N_{Vi}/N_i)$

Binomial-Beta Model (King, Rosen, & Tanner 1999)

- Let's build a Binomial-Beta model!
  - What do we know from the data?

$$T_i = X_i\beta_i^b + (1 - X_i)\beta_i^w \text{ (Accounting identity)}$$

$$\rightsquigarrow \beta_i^w = \left(\frac{T_i}{1-X_i}\right) + \left(\frac{X_i}{1-X_i}\right)\beta_i^b \text{ (Linear function)}$$

  - Let us model the number of voters who turned out $N_{Vi} \in [0, N_i]$

$$N_{Vi} \sim \text{Binomial}(\phi_i, N_i)$$

stochastic component: $p(N_{Vi}|\phi_i, N_i) = \binom{N_i}{N_{Vi}} \phi_i^{N_{Vi}}(1-\phi_i)^{N_i - N_{Vi}}$

systematic component: $\phi_i = X_i\beta_i^b + (1 - X_i)\beta_i^w$, $\mathbb{E}[N_{Vi}] = \phi_i N_i$

$\rightsquigarrow$ Looks like random-coefficient models (with no intercept)!

Binomial-Beta Model (King, Rosen, & Tanner 1999)

- Let us model the number of voters who turned out $N_{Vi} \in [0, N_i]$

$N_{Vi} \sim \text{Binomial}(\phi_i, N_i)$
Stochastic component: $p(N_{Vi}|\phi_i, N_i) = \binom{N_i}{N_{Vi}} \phi_i^{N_{Vi}}(1-\phi_i)^{N_i - N_{Vi}}$
systematic component: $\phi_i = X_i\beta_i^b + (1-X_i)\beta_i^w$, $\mathbb{E}[N_{Vi}] = \phi_i N_i$

$\beta_i^b \sim \text{Beta}(c_b, d_b)$
Stochastic component: $p(\beta_i^b|c_b, d_b) = \frac{\Gamma(c_b+d_b)}{\Gamma(c_b)\Gamma(d_b)}(\beta_i^b)^{c_b-1}(1-\beta_i^b)^{d_b-1}$

$c_b \sim \text{Exp}(\lambda)$
Stochastic component: $p(c_b|\lambda) = \lambda \exp(-\lambda c_b)$
$d_b \sim \text{Exp}(\lambda)$
Stochastic component: $p(d_b|\lambda) = \lambda \exp(-\lambda d_b)$

$\rightsquigarrow$ Equivalent specification for β_i^w, c_w, d_w

Binomial-Beta Model (King, Rosen, & Tanner 1999)

- Writing our posterior distribution

  By Bayes' rule $\cdots$

  - Posterior $\propto$ Likelihood $\times$ Prior
  - $p(\text{parameters}|\text{data}) \propto p(\text{data}|\text{parameters}) \times p(\text{parameters})$

  - $p((\beta_i^b, \beta_i^w)|N_{Vi}, N_i, X_i, i = 1, \ldots, p) \propto$

    $p(N_{Vi}|(\beta_i^b, \beta_i^w), N_i, X_i) \times p((\beta_i^b, \beta_i^w)|c_b, d_b, c_w, d_w) \times p(c_b, d_b, c_w, d_w)$

$$\propto \prod_{i=1}^{p}(X_i\beta_i^b + (1 - X_i)\beta_i^w)^{N_{Vi}}(1 - X_i\beta_i^b - (1 - X_i)\beta_i^w)^{(N_i - N_{Vi})}$$
$$\times \prod_{i=1}^{p} \frac{\Gamma(c_b+d_b)}{\Gamma(c_b)\Gamma(d_b)}(\beta_i^b)^{c_b-1}(1 - \beta_i^b)^{d_b-1}$$
$$\times \prod_{i=1}^{p} \frac{\Gamma(c_w+d_w)}{\Gamma(c_w)\Gamma(d_w)}(\beta_i^w)^{c_w-1}(1 - \beta_i^w)^{d_w-1}$$
$$\times \exp(-\lambda c_b) \times \exp(-\lambda d_b) \times \exp(-\lambda c_w) \times \exp(-\lambda d_w)$$

Binomial-Beta Model (King, Rosen, & Tanner 1999)

- Contextual Effects and Covariates
  - So far, we've assumed that (β_i^b, β_i^w) and X_i are a priori independent
  - $\rightsquigarrow$ black and white turnouts do not depend on % black voters

But...empirical evidence for contextual effects

- Black turnout could be higher in majority-black districts (Fraga 2016, 2018; Oberholzer-Gee & Waldfogel 2001))
- White turnout could be lower in majority-black districts (Gay 2001; Liu 2001)

Incorporate covariate(s) Z_i

$$\phi_i^{Z_i} = X_i \beta_i^{b(Z_i)} + (1 - X_i)\beta_i^{w(Z_i)}$$

$$\beta_i^{b(Z_i)} \sim \text{Beta}(d_b \exp(\alpha + \beta Z_i), d_b);\ \beta_i^{w(Z_i)} \sim \text{Beta}(d_w \exp(\gamma + \delta Z_i), d_w)$$

$$\mathbb{E}[\beta_i^{b(Z_i)}] = \frac{d_b \exp(\alpha+\beta Z_i)}{d_b + d_b \exp(\alpha+\beta Z_i)} = \frac{\exp(\alpha+\beta Z_i)}{1+\exp(\alpha+\beta Z_i)};\ \mathbb{E}[\beta_i^{w(Z_i)}] = \frac{\exp(\gamma+\delta Z_i)}{1+\exp(\gamma+\delta Z_i)}$$

$\rightsquigarrow$ Flat prior on $\alpha, \beta, \gamma, \delta$

Extensions

- Multinomial-Dirichlet models (Rosen, Jiang, King, & Tanner 2001)
- Modeling counts directly (Greiner & Quinn 2009, 2010)
- Temporal dependency (Lewis 2004; Quinn 2004)
- Spatial autocorrelation (Calvo & Escolar 2003)
- Causal inference (Spenkuch 2018; Corvalan, Melo, Shermanandmatt, & Shum 2016)

Classic approaches (Somewhat backwardly)

- Method of Bounds (Duncan & Davis 1953)
  - [0,1] bounds: can't have 100 black voters when 50 people turned out
  - $\max(0, \frac{T_i-(1-X_i)}{X_i}) \leq \beta_i^b \leq \min(\frac{T_i}{X_i}, 1)$
  - $\max(0, \frac{T_i-X_i}{1-X_i}) \leq \beta_i^w \leq \min(\frac{T_i}{1-X_i}, 1)$
- Goodman's (Ecological) Regression (Goodman 1953, 1959)
  - Extreme constancy assumption: $\beta_i^b = B^b; \beta_i^w = B^w$
  - $\mathbb{E}[T_i|X_i] = B^b X_i + B^w(1-X_i)$ (OLS)
- Linear Neighborhood Model (Freedman et al. 1991)
  - No racial difference, but district composition: $\beta_i^b = \beta_i^w = \alpha^c + \alpha^s X_i$
  - $\mathbb{E}[T_i|X_i] = (\alpha^s + \alpha^c)X_i + \alpha^c(1-X_i)$ (OLS)
- Truncated-Normal Model (King's EI) (King 1997)
  - Modeling proportions in the unit square
  - $p(\beta_i^b, \beta_i^w) = \mathrm{TN}(\beta_i^b, \beta_i^w | \mathfrak{B}, \Sigma)$
  - $T_i = \mathfrak{B}^b X_i + \mathfrak{B}^w(1-X_i) + \epsilon_i$
  - $\mathbb{E}[T_i|X_i] = \mathfrak{B}^b X_i + \mathfrak{B}^w(1-X_i)$

Caveats and Future Challenges

- Assumption, Assumption, Assumption!

> *Because some individual-level information is lost in the aggregation process, any single approach to the ecological inference problem will by necessity require a set of modeling assumptions, and* **the success of the endeavor will depend on these assumptions***. It is therefore of value to the data analyst to have a variety of models with which to explore the data (King, Rosen, & Tanner 1999, 62).*

- Computational burden
  - Convergence issues
  - Cross-sectional and time-series EI

⇝ **Let's estimate Binomial-Beta models with JAGS!**