

## 0.1 ei.RxC: Hierarchical Multinomial-Dirichlet Ecological Inference Model for $R \times C$ Tables

Given  $n$  contingency tables, each with observed marginals (column and row totals), ecological inference (EI) estimates the internal cell values in each table. The hierarchical Multinomial-Dirichlet model estimates cell counts in  $R \times C$  tables. The model is implemented using a nonlinear least squares approximation and, with bootstrapping for standard errors, had good frequentist properties.

### Syntax

```
> z.out <- zelig(cbind(T0, T1, T2, T3) ~ X0 + X1,
                 covar = NULL,
                 model = "ei.RxC", data = mydata)
> x.out <- setx(z.out, fn = NULL)
> s.out <- sim(z.out)
```

### Inputs

- $T_0, T_1, T_2, \dots, T_C$ : numeric vectors (either counts, or proportions that sum to one for each row) containing the column margins of the units to be analyzed.
- $X_0, X_1, X_2, \dots, X_R$ : numeric vectors (either counts, or proportions that sum to one for each row) containing the row margins of the units to be analyzed.
- **covar**: (optional) a covariate that varies across tables, specified as `covar = ~ Z1`, for example. (The model only accepts one covariate.)

### Examples

1. Basic examples: No covariate  
Attaching the example dataset:

```
> data(Weimar)
```

Estimating the model:

```
> z.out <- zelig(cbind(Nazi, Government, Communists, FarRight,
+   Other) ~ shareunemployed + shareblue + sharewhite + shareself +
+   sharedomestic, model = "ei.RxC", data = Weimar)
> summary(z.out)
```

Estimate fractions of different social groups that support political parties:

```
> s.out <- sim(z.out, num = 10)
```

Summarizing fractions of different social groups that support political parties:

```
> summary(s.out)
```

## 2. Example of covariates being present in the model

Using the example dataset Weimar and estimating the model

```
> z.out <- zelig(cbind(Nazi, Government, Communists, FarRight,  
+   Other) ~ shareunemployed + shareblue + sharewhite + shareself +  
+   sharedomestic, covar = ~shareprotestants, model = "ei.RxC",  
+   data = Weimar)  
> summary(z.out)
```

Set the covariate to its default (mean/median) value

```
> x.out <- setx(z.out)
```

Estimate fractions of different social groups that support political parties:

```
> s.out <- sim(z.out, num = 100)
```

Summarizing fractions of different social groups that support political parties:

```
> s.out <- summary(s.out)
```

## Model

Consider the following  $5 \times 5$  contingency table for the voting patterns in Weimar Germany. For each geographical unit  $i$  ( $i = 1, \dots, p$ ), the marginals  $T_{1i}, \dots, T_{Ci}$ ,  $X_{1i}, \dots, X_{Ri}$  are known for each of the  $p$  electoral precincts, and we would like to estimate  $(\beta_i^r, r = 1, \dots, R, c = 1, \dots, C - 1)$  which are the fractions of people in social class  $r$  who vote for party  $c$ , for all  $r$  and  $c$ .

|            | Nazi           | Government     | Communists     | Far Right      | Other                           |         |
|------------|----------------|----------------|----------------|----------------|---------------------------------|---------|
| Unemployed | $\beta_{11}^i$ | $\beta_{12}^i$ | $\beta_{13}^i$ | $\beta_{14}^i$ | $1 - \sum_{c=1}^4 \beta_{1c}^i$ | $X_1^i$ |
| Blue       | $\beta_{21}^i$ | $\beta_{22}^i$ | $\beta_{23}^i$ | $\beta_{24}^i$ | $1 - \sum_{c=1}^4 \beta_{2c}^i$ | $X_2^i$ |
| White      | $\beta_{31}^i$ | $\beta_{32}^i$ | $\beta_{33}^i$ | $\beta_{34}^i$ | $1 - \sum_{c=1}^4 \beta_{3c}^i$ | $X_3^i$ |
| Self       | $\beta_{41}^i$ | $\beta_{42}^i$ | $\beta_{43}^i$ | $\beta_{44}^i$ | $1 - \sum_{c=1}^4 \beta_{4c}^i$ | $X_4^i$ |
| Domestic   | $\beta_{51}^i$ | $\beta_{52}^i$ | $\beta_{53}^i$ | $\beta_{54}^i$ | $1 - \sum_{c=1}^4 \beta_{5c}^i$ | $X_5^i$ |
|            | $T_{1i}$       | $T_{2i}$       | $T_{3i}$       | $T_{4i}$       | $1 - \sum_{c=1}^4 \beta_{ci}$   |         |

The marginal values  $X_{1i}, \dots, X_{Ri}$ ,  $T_{1i}, \dots, T_{Ci}$  may be observed as counts or fractions.

Let  $T'_i = (T'_{1i}, T'_{2i}, \dots, T'_{Ci})$  be the number of voting age persons who turn out to vote for different parties. There are three levels of hierarchy in the Multinomial-Dirichlet EI model. At the first stage, we model the data as:

- The *stochastic component* is described  $T'_i$  which follows a multinomial distribution:

$$T'_i \sim \text{Multinomial}(\Theta_{1i}, \dots, \Theta_{Ci})$$

- The *systematic components* are

$$\Theta_{ci} = \sum_{r=1}^R \beta_{rc}^i X_{ri} \quad \text{for } c = 1, \dots, C$$

At the second stage, we use an optional covariate to model  $\Theta_{ci}$ 's and  $\beta_{rc}^i$ :

- The *stochastic component* is described by  $\beta_r^i = (\beta_{r1}, \beta_{r2}, \dots, \beta_{r,C-1})$  for  $i = 1, \dots, p$  and  $r = 1, \dots, R$ , which follows a Dirichlet distribution:

$$\beta_r^i \sim \text{Dirichlet}(\alpha_{r1}^i, \dots, \alpha_{rC}^i)$$

- The *systematic components* are

$$\alpha_{rc}^i = \frac{d_r \exp(\gamma_{rc} + \delta_{rc} Z_i)}{d_r (1 + \sum_{j=1}^{C-1} \exp(\gamma_{rj} + \delta_{rj} Z_i))} = \frac{\exp(\gamma_{rc} + \delta_{rc} Z_i)}{1 + \sum_{j=1}^{C-1} \exp(\gamma_{rj} + \delta_{rj} Z_i)}$$

for  $i = 1, \dots, p$ ,  $r = 1, \dots, R$ , and  $c = 1, \dots, C - 1$ .

In the third stage, we assume that the regression parameters (the  $\gamma_{rc}$ 's and  $\delta_{rc}$ 's) are *a priori* independent, and put a flat prior on these regression parameters. The parameters  $d_r$  for  $r = 1, \dots, R$  are assumed to follow exponential distributions with mean  $\frac{1}{\lambda}$ .

## Output Values

The output of each Zelig command contains useful information which you may view. For example, if you run

```
> z.out <- zelig(cbind(T0, T1, T2) ~ X0 + X1 + X2,
  model = "ei.RxC", data = mydata)
```

then you may examine the available information in `z.out` by using `names(z.out)`. For example,

- From the `zelig()` output object `z.out$coefficients` are the estimates of  $\gamma_{ij}$  (and also  $\delta_{ij}$ , if covariates are present). The parameters are returned as a single vector of length  $R \times (C - 1)$ . If there is a covariate,  $\delta$  is concatenated to it.
- From the `sim()` output object, you may extract the parameters  $\beta_{ij}$  corresponding to the estimated fractions of different social groups that support different political parties, by using `s.out$qi$ev`. For each precinct, that will be a matrix with dimensions: simulations  $\times R \times C$ .  
`summary(s.out)` will give you the nationwide aggregate parameters.

## How to Cite

To cite the *ei.RxC* Zelig model use:

Jason Wittenberg, Ferdinand Alimadhi, Badri Narayan Bhaskar, and Olivia Lau. 2007. “ei.RxC: Hierarchical Multinomial-Dirichlet Ecological Inference Model,” in Kosuke Imai, Gary King, and Olivia Lau, “Zelig: Everyone’s Statistical Software,” <http://gking.harvard.edu/zelig>.

To cite Zelig as a whole, please reference these two sources:

Kosuke Imai, Gary King, and Olivia Lau. 2007. “Zelig: Everyone’s Statistical Software,” <http://GKing.harvard.edu/zelig>.

Kosuke Imai, Gary King, and Olivia Lau. 2008. “Toward A Common Framework for Statistical Analysis and Development,” *Journal of Computational and Graphical Statistics*, forthcoming, <http://gking.harvard.edu/files/abs/z-abs.shtml>.

## See also

For more information please see Rosen et al. (2001)

# Bibliography

Rosen, O., Jiang, W., King, G., and Tanner, M. A. (2001), “Bayesian and Frequentist Inference for Ecological Inference: The  $R \times C$  Case,” *Statistica Neerlandica*, 55, 134–156, <http://gking.harvard.edu/files/abs/rosen-abs.shtml>.