

MODELIZACIÓN ONOMÁSTICA

María José Ginzo-Villamayor

Departamento de Estadística e Investigación Operativa
Universidade de Santiago de Compostela



DEPARTAMENTO DE ESTATÍSTICA
E INVESTIGACIÓN OPERATIVA



INSTITUTO DE MATEMÁTICAS

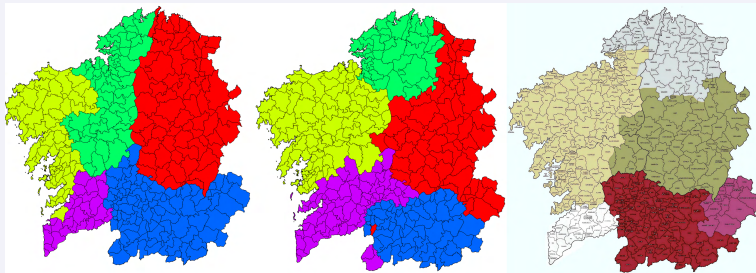
Motivación

- ▶ Os apelidos poden empregarse como **fonte de información para caracterizar poboacións** xa que a análise do patrón dun apelido proporciona información sobre tendencias de movementos de poboación ao longo do tempo.

Obxectivo

- ▶ A través de diferentes **medidas de isonimia** entre rexións poderanse identificar patróns de apelidos, construíndo *clusters* de zonas de apelidos, é dicir, obter unha representación gráfica das diferentes rexións de apelidos.

Cluster de apelidos para a distancia Lasker (5 grupos)



Actualidade — Nacidos e 1965 ou antes — Dióceses de Galicia

- ▶ Isonimia: $I_i = \sum_k p_{ki}^2$, onde p_{ki} é a frecuencia relativa dun apelido k en unha rexión i .
- ▶ Isonimia entre dúas rexións i e j : $I_{ij} = \sum_k p_{ki} p_{kj}$
- ▶ Distancia de Lasker: $L = -\log(I_{ij})$



Ginzo-Villamayor, M.J.; Crujeiras, R. M. e Sousa Fernández, X. (2013) Surname patterns in Galicia.

Libro de Actas do congreso. XI Congreso Galego de Estatística e Investigación de Operacións. A Coruña (España).

- ▶ As diferentes liñas de investigación no contexto da onomástica non soen ter en conta a **dimensión espacial e espazo-temporal da evolución** dos apelidos.
- ▶ Fixando **rexións administrativas** (e.g. concellos) pode facerse uso de métodos espaciais e espazo-temporais para a análise de datos de conteo que permitan modelar o patrón subxacente á evolución dos apelidos.
- ▶ Para analizar o patrón espacial dos apelidos en Galicia axustouse o modelo proposto por Besag et al. (1991) (**modelo BYM**), adaptado a este contexto.
- ▶ Para o axuste de modelos xerárquicos neste contexto faise uso da metodoloxía INLA (**Integrated Nested Laplace Approximation**) proposta por Rue et al. (2009).
- ▶ Datos proporcionados polo Instituto Galego de Estatística.



Besag, J., York, J. and Mollié, A. (1991) Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.



Rue, H., Martino, S. and Chopin, N (2009) Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.



Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall, CRC Press, London.

Modelos Bayesianos espaciais

- Fixemos un apelido ...

Y_i = número de persoas con ese apelido no concello i , $i = 1, \dots, n$
este proceso de conteo será modelado a través dun modelo Poisson–LogNormal,

$$Y_i | \eta_i \sim \text{Pois}(E_i \exp(\eta_i)),$$

onde E_i é a *poboación en risco*, η_i un predictor lineal e as variables Y_i son independentes.

$$\eta_i = \mu + f_s(s_i) + f_u(s_i),$$

onde s_i é o centroide de cada concello, f_s e f_u son os efectos estruturados e non estruturados. f_s , é un campo aleatorio Gaussiano de Markov. f_u , ruído branco que representa a “sobredispersión” dos concellos.



Banerjee S., Carlin, B. E. e Gelfand, A. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability



Rue, H. e Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall, CRC Press, London.

Axuste e estimación

- ▶ INLA propociona unha ferramenta rápida e útil para axustar modelos gaussianos latentes (os procesos que rixen f_s , f_u e f_T teñen distribucións Gaussianas), incluíndo modelos con estrutura temporal ou espacial nun contexto Bayesiano.
- ▶ O procedemento INLA calcula a aproximación numérica das distribucións a posteriori, baseados no método de *aproximación de Laplace*.
- ▶ Hai que especificar os hiperparámetros do efecto espacial a priori e da parte aleatoria.

```
> prior.spat=c(1,0.00005)
```

```
> hyper.spat=list(prec=list(param=prior.spat))
```

```
> prior.iid=c(1,0.00005)
```

```
> hyper.iid=list(prec=list(param=prior.iid))
```

- ▶ Para a análise de sensibilidade do modelo empregouse o criterio **Criterio de Información de Desviación (DIC)**.

Apelidos Crujeiras, Ginzo e Rodríguez (Padrón 2011)

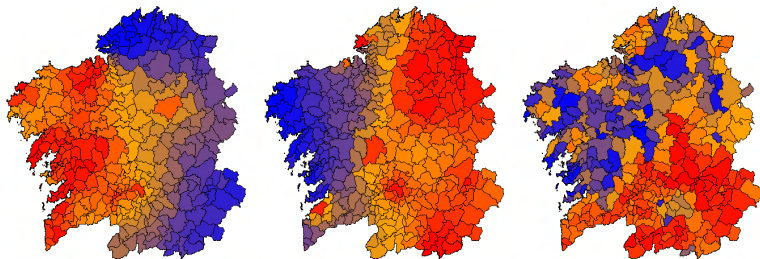


Figura: Media a posteriori da parte estruturada do campo latente.

← Crujeiras — Ginzo — Rodríguez

Apelidos Crujeiras, Ginzo e Rodríguez (Padrón 2011)

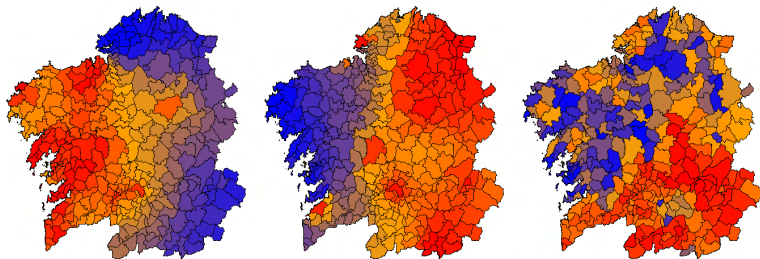


Figura: Media a posteriori da parte estruturada do campo latente.

← Crujeiras — Ginzo — Rodríguez

Apelidos Crujeiras, Ginzo e Rodríguez (Padrón 2011)

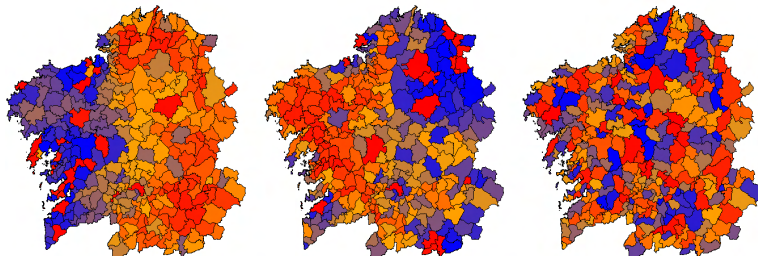


Figura: Media a posteriori do efecto non estruturado do campo latente.

Máis exemplos

▶ Liñas futuras

Máis exemplos

▶ Liñas futuras

Modelos Bayesianos espazo-temporais

- ▶ Fixemos un apelido e un instante de tempo ...

Y_{it} = número de persoas con ese apelido no concello i no tempo t
que será observada nos n concellos e en T instantes do tempo

$$Y_{it} | \eta_i \sim \text{Pois}(E_{it} \exp(\eta_{it})).$$

O modelo espacial esténdese ao permitir a compoñente temporal quedando:

$$\eta_{it} = \mu + f_s(s_i) + f_u(s_i) + f_T(t), t = 1, \dots, T.$$

onde $f_T(t)$ é a estrutura temporal.

Apellido Ginzo en Galicia (2000-2015)

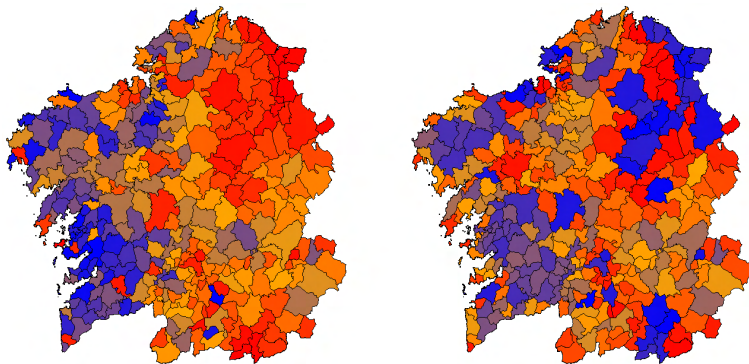


Figura: Esquerda: Media a posteriori do efecto espacial principal.
Dereita: Efecto debido ao tempo.

- └ Unha mirada máis profunda en certos apelidos
- └ Modelos espaciais e espazo-temporais



Ginzo-Villamayor, M.J. and Crujeiras, R. (2015)

Modelización Onomástica.

Libro de actas: ACTAS XII Congreso Galego de Estatística e Investigación de Operacións Lugo, 22-23-24 de outubro de 2015. ISBN 978-84-8192-522-7. (pp: 292-299). XII Congreso Galego de Estatística e Investigación de Operacións. Lugo (España).

Aínda queda traballo por facer ...

- ▶ Estender o traballo a outros apelidos e nomes para completar o mapa de apelidos/nomes en Galicia. [◀ Máis exemplos](#)
- ▶ Estender o período de tempo no estudo.
- ▶ Investigar o uso dalgunha covariable nestes modelos.

Aínda queda traballo por facer ...

- ▶ Estender o traballo a outros apelidos e nomes para completar o mapa de apelidos/nomes en Galicia. [◀ Máis exemplos](#)
- ▶ Estender o período de tempo no estudo.
- ▶ Investigar o uso dalgunha covariable nestes modelos.

MODELIZACIÓN ONOMÁSTICA

María José Ginzo-Villamayor

Departamento de Estadística e Investigación Operativa
Universidade de Santiago de Compostela



DEPARTAMENTO DE ESTATÍSTICA
E INVESTIGACIÓN OPERATIVA



INSTITUTO DE MATEMÁTICAS