New Results from Multilevel Models of the Speech Community

Josef Fruehwald
Laurel MacKenzie

University of Pennsylvania

October 28, 2011
NWAV 40

Outline

Outline

Introduction

Statistics and Sociolinguists

- It has been recently argued that hierarchical statistical models are statistically appropriate for sociolinguistic data (Gorman, 2009; Johnson, 2009).
- Hierarchical models are also *conceptually appropriate*, since they model variation in the way sociolinguists view the world.

Sociolinguistics in Statistics

Different variables exhibit different rates of “community cohesion” in ways which are potentially interesting.

Outline

Flat Models

Historically. . .

- In the past, all models were “flat,” pooling all speakers’ data together.
- This was largely due to mathematical and computational limitations.

Community Grammar

- Perhaps this is a feature, not a bug, as the object of sociolinguistic inquiry is larger than one speaker.
- But the statistics “see” the community grammar more like the Borg Collective.

Hierarchical Models

Now…

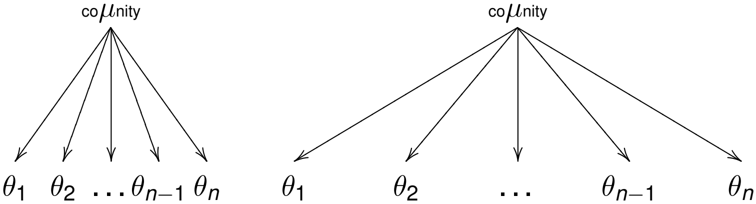
- More work is emphasizing the statistical appropriateness of hierarchical models for sociolinguistic data.

Community Grammar

- Hierarchical models also “see” variation as sociolinguists do: individuals coordinating themselves around a communal norm. (Labov, 2010)

Hierarchical Models

- μ = Communal norm for some variable.
- θ_i = An individual speaker's rate for some variable.



- κ will describe the degree of κohesion around the coμnal norm, with higher κ values indicating greater cohesion.

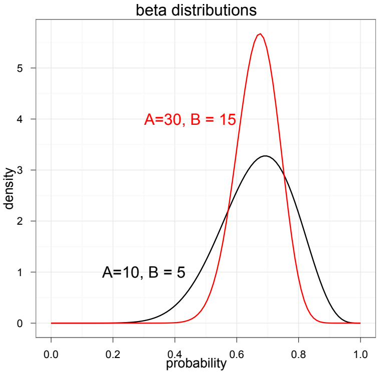
Hierarchical Models

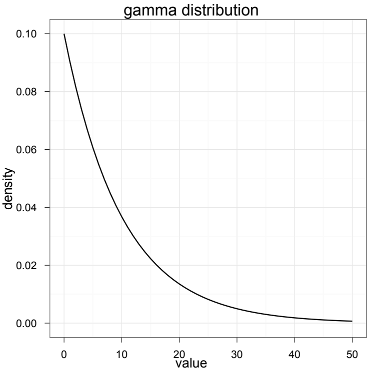
A mathematical definition

- $\mu \sim$ beta(A, B)
- $\theta_i \sim$ beta($\kappa \times \mu$, $\kappa \times (1 - \mu)$)
- $\{A, B, \kappa\} \sim$ gamma(1, $\frac{1}{10}$)

- A beta distribution describes probabilities over probabilities.
  - A>B, skewed towards 1
  - A<B, skewed towards 0
  - As A+B increases, so does the peakedness of the distribution.
- This gamma distribution is skewed towards small numbers, and is the common sort of distribution used to describe parameters like A, B and κ.

Hierarchical Models





Hierarchical Models

Markov Chain Monte Carlo Estimation

MCMC is a form of simulation which iteratively updates model parameters based on:

1. The value this parameter had in the last iteration of the simulation.
2. The likelihood of some value given the data, and the other parameter values.
3. Some prior expectations about likely values for this parameter.

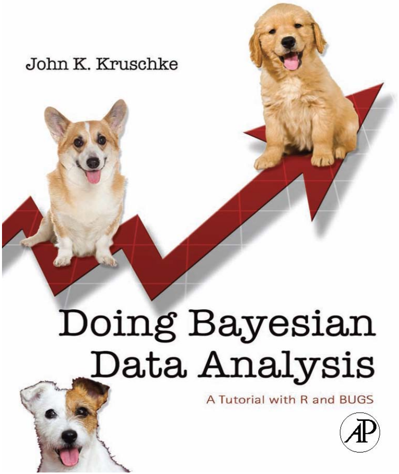
Flexibility

It allows you to define models which do not submit to a regression formulation.

Hierarchical Models

Markov Chain Monte Carlo Estimation

Further Reading



Implementation

- JAGS (just another Gibbs sampler).
- `rjags` package in R.

Outline

Case Studies

All of our models will involve:

- One or more μ,
- speaker level θ based on μ,
- a community cohesion coefficient κ for each μ,
- the inclusion of additional predictors relevant to the variable at hand.

TD Retention

Regular Past vs. Semiweak

- Regular past $>$ Semiweak (Labov et al., 1968)
  - *packed* $>$ *kept*
- Guy (1991): Difference is caused by one process with probability *p* applying once to regular past, and twice to semiweak.
- Bermúdez-Otero (2010), Fruehwald (2011): There are two independent processes. Process one applies to both regular and semiweak, process two applies only to semiweak.

TD Retention

Data

- Data is drawn from the Buckeye Corpus (Pitt et al., 2005).
    - 38 speakers from in or around Columbus, OH.
- 1,677 tokens of regular past.
- 329 tokens of semiweak past.

TD Retention

Model

Parameters

- One community-level μ for each process: μ_{one} and μ_{two}
- One cohesion coefficient related to each μ: κ_{one} and κ_{two}
- Two individual level parameters per speaker: θ_{one} and θ_{two}

Regression

- For each speaker, probability of retention = $\hat{\theta}$.
  - $\hat{\theta} = \theta_{one}$ for past tense.
  - $\hat{\theta} = \theta_{one} \times \theta_{two}$ for semiweak.
- $\hat{\theta}$ is scaled by additive logistic effects associated with preceding and following segments.

TD Retention

Results

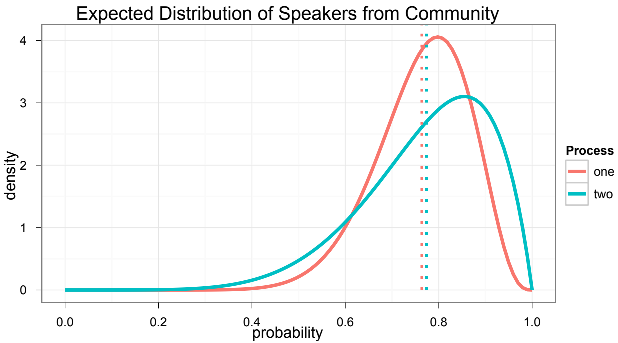
	Process One	Process Two
μ	0.76	0.77
A + B	22.08	20.89
κ	17.56	8.74

Preceding Segment	Effect
/r/	1.38
/l/	0.39
fricative	0.11
stop	−0.18
sibilant	−0.65
nasal	−1.06

Following Segment	Effect
vowel	1.19
pause	1.02
/y/	0.27
/n/	-0.43
stop	-0.49
/l/	-0.53
sibilant	-0.82

TD Retention

Results



- regular past = process one
- semiweak = process one × process two

TD Retention

Conclusion

- Based on the different patterns in the speech community, it appears that there are two different processes involved in TD Retention.

Auxiliary Contraction

- e.g. *He* $\underline{\textit{has}}$ $\sim$ *he*$\underline{\textit{'s}}$ *arrived*
- $$\left\{ \begin{array}{l} \text{NP} \\ \text{pronoun} \end{array} \right\} \left\{ \begin{array}{l} \textit{had} \\ \textit{has} \\ \textit{have} \end{array} \right\}$$
- Is contraction. . .
  - . . . a single, pan-auxiliary process?
  - . . . multiple independent auxiliary-specific processes?

Auxiliary Contraction

What underlies the surface alternation?

- Variable adjunction of an auxiliary to the element that precedes it (Kaisse 1983)

  $\text{X}^\frown\text{Aux} \rightsquigarrow [[\text{X}]\text{Aux}]$

- Each auxiliary has a full and a short allomorph; short form conditioned when adjunction has occurred

full	/hæd/	/hæz/	/hæv/
short	/d/	/z/	/v/

e.g. $\text{3.sg.fem}^\frown\text{T}_{\text{[past]}} \rightsquigarrow [[\text{3.sg.fem}]\text{T}_{\text{[past]}}] \rightarrow [\text{ʃid}]$
$\quad\;\; \text{3.sg.fem}^\frown\text{T}_{\text{[past]}} \rightsquigarrow \text{3.sg.fem}^\frown\text{T}_{\text{[past]}} \rightarrow [\text{ʃi hæd}]$

Auxiliary Contraction

Data

- Data drawn from the Philadelphia Neighborhood Corpus (Labov & Rosenfelder 2011)
- 39 speakers from Philadelphia.

had	has	have
105	147	263

Auxiliary Contraction

Model

Parameters

- One μ for each auxiliary (μ_{had}, μ_{has}, μ_{have})

- One κ for all three auxiliaries (κ_{same})
- One κ for each auxiliary (κ_{diff})

- A parameter which selects a κ_{diff} or κ_{same} preference
- For each speaker, θ_{had}, θ_{has}, θ_{have}

Regression

- Additive logit effects for pronoun vs NP subjects (MacKenzie 2010)
- If the subject is an NP, slope for how many words in the NP (MacKenzie 2011)

Auxiliary Contraction

Results

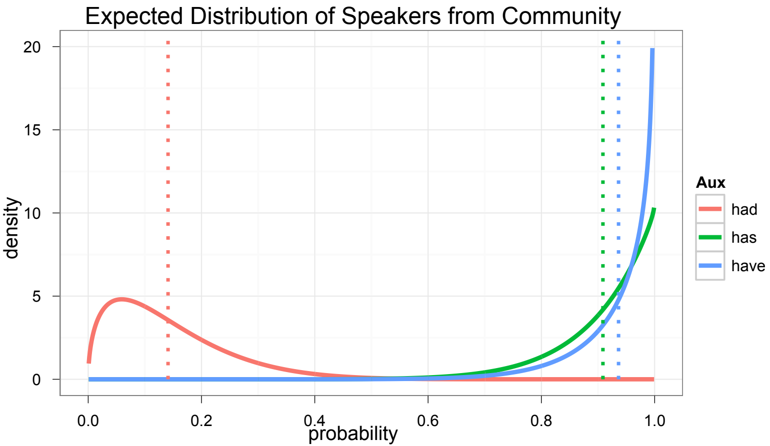
	had	has	have
μ	0.14	0.91	0.94
A + B	19.82	20.72	19.44
κ_{diff}	10.89	8.48	11.61
κ_{same}		10.78	

κ_{same} preference = 0.53 ($>$ .5 favors)

	Effect
NP	-1.29
NP Length +1	-2.59

Auxiliary Contraction

Results



Auxiliary Contraction

Conclusion

- The similar patterns for each auxiliary in the speech community support the proposal that a single process underlies contraction of all auxiliaries.

Outline

Conclusions

Case Studies

TD Retention

Based on the different patterns in the speech community, it appears that there are two, different processes involved in TD Retention.

Auxiliary Contraction

The different patterns in the speech community support the proposal that a single process underlies contraction of all auxiliaries.

Conclusions

Statistics

Hierarchical Models

- Hierarchical models are conceptually appropriate for sociolinguistic data.
- Hierarchical models can give us parameters that other models can't — namely, degree of community cohesion — which can provide support for linguistic reasoning.

MCMC

- Markov Chain Monte Carlo methods allow us to fit and compare interesting models which cannot be fit as ordinary regression.

Further Directions

Modeling Change

Changing Communal Norms

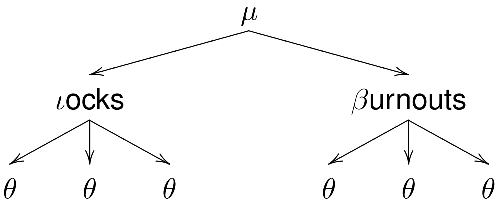
- $\mu \sim$ time

Changing Cohesion

- $\kappa \sim$ time

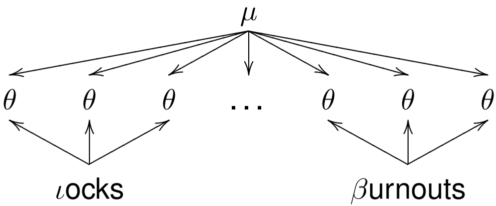
Further Directions

Modeling Social Factors



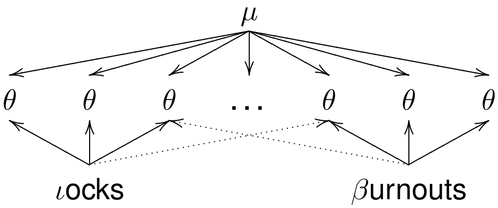
Further Directions

Modeling Social Factors



Further Directions

Modeling Social Factors



Acknowledgments

Special thanks to...

Bill Labov, Gillian Sankoff, Tony Kroch, Charles Yang, Michael Friesner, Ingrid Rosenfelder, Toni Cook, Caitlin Light, Hilary Prichard, Meredith Tamminga, Danielle Turton, and Sabriya Fisher for their thoughtful comments and criticisms.

Supported by

The Benjamin Franklin Fellowship

References

- Bermúdez-Otero, Ricardo. 2010. “Currently available data on English t/d-deletion fail to refute the classical modular feedforward architecture of phonology”. Paper presented at MFM 18.
- Fruehwald, Josef. 2011. “Redevelopment of a Morphological Class”. Paper presented at PLC 35.
- Gorman, Kyle. 2009. “Hierarchical regression modeling for language research“. University of Pennsylvania Institute for Research in Cognitive Science Technical Report. URL `http://repository.upenn.edu/ircs_reports/202/`.
- Johnson, Daniel Ezra. 2009.“Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis.” *Language and Linguistic Compass* 3:359383.
- Kaisse, Ellen M. 1983. “The syntax of auxiliary reduction in English.” *Language* 59:93122.
- Labov, William. 2010. “What is to be learned?” Paper presented at NELS 41.
- Labov, William, and Ingrid Rosenfelder. 2011. *The Philadelphia Neighborhood Corpus.*
- MacKenzie, Laurel. 2010. “Locating linguistic variation: A case study of English auxiliary contraction.” Poster presented at NELS 41, University of Pennsylvania.
- MacKenzie, Laurel. 2011. “English auxiliary contraction as a two-stage process: Evidence from corpus data.” Paper presented at WCCFL 29, University of Arizona.
- Pitt, M. A., L. Dilley, Keith Johnson, S. Kiesling, W. Raymond, Elizabeth Hume, and E. Fosler-Lussier. 2007. *Buckeye corpus of conversational speech (2nd release)*. URL `www.buckeyecorpus.osu.edu`.