Collecting sociolinguistic micro-judgments of acoustic cues online

Kathryn Campbell-Kipler, Kiwako Ito, Eric Ruppe, Elizabeth McCullough, Mary E. Beckman, Tsz-Him Tsui

http://ling.osu.edu/research/groups/soundsize
Department of Linguistics
The Ohio State University

Our goal

Reliably collect and analyze micro-judgments of sociolinguistic perception, to align with specific acoustic features.

Our goal

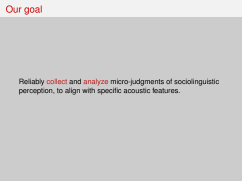
Reliably collect and analyze micro-judgments of sociolinguistic perception, to align with specific acoustic features.

This tutorial is about getting and analyzing sociolinguistic perception data. The projects I'm going to talk about are focused on what I'm calling micro-judgments, where participants get a very small amount of linguistic material ad make one relatively simple judgment. The issues that we're dealing with, though, are liking to apply to more complex evaluative tasks.

I'll be talking about four different studies, in different stages of completion, which use or will use these methods. I won't get into too much detail on each one, but instead will be focusing on two of the common dimensions: tools for collecting and tools for analyzing these kinds of data.

Our goal

Reliably collect and analyze micro-judgments of sociolinguistic perception, to align with specific acoustic features.



This tutorial is about getting and analyzing sociolinguistic perception data. The projects I'm going to talk about are focused on what I'm calling micro-judgments, where participants get a very small amount of linguistic material ad make one relatively simple judgment. The issues that we're dealing with, though, are liking to apply to more complex evaluative tasks.

I'll be talking about four different studies, in different stages of completion, which use or will use these methods. I won't get into too much detail on each one, but instead will be focusing on two of the common dimensions: tools for collecting and tools for analyzing these kinds of data.

Pin vs. pen

- Listeners can reorganize lexical competition on the fly to reflect speaker variation (Dahan et al., 2008, e.g.)
- Sociolinguistic expectations, triggered by photos, can influence linguistic processing (Hay et al., 2006; Staum Casasanto, 2008)

Pin vs. pen

- Listeners can reorganize lexical competition on the fly to reflect speaker variation (Dahan et al., 2008, e.g.)
- Sociolinguistic expectations, triggered by photos, can influence linguistic processing (Hay et al., 2006; Staum Casasanto, 2008)

This project is a collaboration between Kiwako Ito and Kathryn Campbell-Kibler. A new branch of it with Mary Beckman and Liz McCullough, looking at the production data of the eyetracking participants will be discussed tomorrow in Tutorial 3.

Pin vs. pen

- Listeners can reorganize lexical competition on the fly to reflect speaker variation (Dahan et al., 2008, e.g.)
- Sociolinguistic expectations, triggered by photos, can influence linguistic processing (Hay et al., 2006; Staum Casasanto, 2008)

Do sociolinguistic expectations regarding race and/or class shift patterns of eye-tracking based on the pin/pen merger? If so, does adaptation based on direct exposure to speech eliminate or mitigate that effect?

Pin vs. pen

- Listeners can reorganize lexical competition on the fly to reflect speaker variation (Dahan et al., 2008, e.g.)
- Sociolinguistic expectations, triggered by photos, can influence linguistic processing (Hay et al., 2006; Staum Casasanto, 2008)

Do sociolinguistic expectations regarding race and/or class shift patterns of eye-tracking based on the pin/pen merger? If so, does adaptation based on direct exposure to speech eliminate or mitigate that effect?

This project is a collaboration between Kiwako Ito and Kathryn Campbell-Kibler. A new branch of it with Mary Beckman and Liz McCullough, looking at the production data of the eyetracking participants will be discussed tomorrow in Tutorial 3.

Pin vs. pen

- Listeners can reorganize lexical competition on the fly to reflect speaker variation (Dahan et al., 2008, e.g.)
- Sociolinguistic expectations, triggered by photos, can influence linguistic processing (Hay et al., 2006; Staum Casasanto, 2008)

Do sociolinguistic expectations regarding race and/or class shift patterns of eye-tracking based on the pin/pen merger? If so, does adaptation based on direct exposure to speech eliminate or mitigate that effect?

Which tokens are more or less ambiguous? Which speakers are more or less merged?

Pin vs. pen

- Listeners can reorganize lexical competition on the fly to reflect speaker variation (Dahan et al., 2008, e.g.)
- Sociolinguistic expectations, triggered by photos, can influence linguistic processing (Hay et al., 2006; Staum Casasanto, 2008)

Do sociolinguistic expectations regarding race and/or class shift patterns of eye-tracking based on the pin/pen merger? If so, does adaptation based on direct exposure to speech eliminate or mitigate that effect?

Which tokens are more or less ambiguous? Which speakers are more or less merged?

This project is a collaboration between Kiwako Ito and Kathryn Campbell-Kibler. A new branch of it with Mary Beckman and Liz McCullough, looking at the production data of the eyetracking participants will be discussed tomorrow in Tutorial 3.

Face-voice connections

- Visual representation of faces integrate with voice-based social information (Campanella & Belin, 2007; de Gelder et al., 2002; Williams et al., 1976)
- Combining tasks which align conscious and automatic processes with those which oppose them (process dissociation) can illuminate their relationship (Jacoby, 1991; Payne, 2005; Payne & Stewart, 2007)

Face-voice connections

- Visual representation of faces integrate with voice-based social information (Campanella & Belin, 2007; de Gelder et al., 2002; Williams et al., 1976)
- Combining tasks which align conscious and automatic processes with those which oppose them (process dissociation) can illuminate their relationship (Jacoby, 1991; Payne, 2005; Payne & Stewart, 2007)

This is work I'm doing on my own (Kathryn Campbell-Kibler) and have just begun. Pilot data on male faces has been collected.

Face-voice connections

- Visual representation of faces integrate with voice-based social information (Campanella & Belin, 2007; de Gelder et al., 2002; Williams et al., 1976)
- Combining tasks which align conscious and automatic processes with those which oppose them (process dissociation) can illuminate their relationship (Jacoby, 1991; Payne, 2005; Payne & Stewart, 2007)

Do social evaluations of faces and voices differ when presented in isolation, combined as a purportedly single person or simply co-presented?

Face-voice connections

- Visual representation of faces integrate with voice-based social information (Campanella & Belin, 2007; de Gelder et al., 2002; Williams et al., 1976)
- Combining tasks which align conscious and automatic processes with those which oppose them (process dissociation) can illuminate their relationship (Jacoby, 1991; Payne, 2005; Payne & Stewart, 2007)

Do social evaluations of faces and voices differ when presented in isolation, combined as a purportedly single person or simply co-presented?

This is work I'm doing on my own (Kathryn Campbell-Kibler) and have just begun. Pilot data on male faces has been collected.

Foreign accent

- Information in the speech signal alone can lead listeners to make social judgments. (Purnell et al., 1999)
- Social judgments regarding non-native speakers are common and often harsh. (Brennan & Brennan, 1981; Gluszek & Dovidio, 2010; Kinzler et al., 2007)

- Information in the speech signal alone can lead listeners to make social judgments. (Purnell et al., 1999)
- Social judgments regarding non-native speakers are common and often harsh. (Brennan & Brennan, 1981; Gluszek & Dovidio, 2010; Kinzler et al., 2007)

This project is Liz McCullough's dissertation.

Foreign accent

- Information in the speech signal alone can lead listeners to make social judgments. (Purnell et al., 1999)
- Social judgments regarding non-native speakers are common and often harsh. (Brennan & Brennan, 1981; Gluszek & Dovidio, 2010; Kinzler et al., 2007)

How do US English native speakers assess degree of foreign accent across multiple L1 backgrounds? What acoustic properties influence their assessments?

Foreign accent

- Information in the speech signal alone can lead listeners to make social judgments. (Purnell et al., 1999)
- Social judgments regarding non-native speakers are common and often harsh. (Brennan & Brennan, 1981; Gluszek & Dovidio, 2010; Kinzler et al., 2007)

How do US English native speakers assess degree of foreign accent across multiple L1 backgrounds? What acoustic properties influence their assessments?

This project is Liz McCullough's dissertation.

Sibilants and gender

Several languages show gendering of /s/ and/or /ʃ/:

- English (Campbell-Kibler, 2011; Crist, 1997; Linville, 1998)
- Danish (Pharao and colleagues, in progress)
- Mandarin (Hu, 1991; Li, 2005)
- Japanese (Chew, 1969)

Sibilants and gender

Several languages show gendering of /s/ and/or /ʃ/:

- English (Campbell-Kibler, 2011; Crist, 1997; Linville, 1998)
- Danish (Pharao and colleagues, in progress)
- Mandarin (Hu, 1991; Li, 2005)
- Japanese (Chew, 1969)

This is a shared project across the Sound Size collective, with Mark Beckman, Kiwako Ito, Kathryn Campbell-Kibler, Tsz-Him Tsui, Eric Ruppe and Liz McCullough. Eric and I have been primarily working on this phase.

Sibilants and gender

Several languages show gendering of /s/ and/or /ʃ/:

- English (Campbell-Kibler, 2011; Crist, 1997; Linville, 1998)
- Danish (Pharao and colleagues, in progress)
- Mandarin (Hu, 1991; Li, 2005)
- Japanese (Chew, 1969) Goes the other way!

Several languages show gendering of /s/ and/or /ʃ/:

- English (Campbell-Kibler, 2011; Crist, 1997; Linville, 1998)
- Danish (Pharao and colleagues, in progress)
- Mandarin (Hu, 1991; Li, 2005)
- Japanese (Chew, 1969) Goes the other way!

This is a shared project across the Sound Size collective, with Mark Beckman, Kiwako Ito, Kathryn Campbell-Kibler, Tsz-Him Tsui, Eric Ruppe and Liz McCullough. Eric and I have been primarily working on this phase.

Sibilants and gender

Several languages show gendering of /s/ and/or /ʃ/:

- English (Campbell-Kibler, 2011; Crist, 1997; Linville, 1998)
- Danish (Pharao and colleagues, in progress)
- Mandarin (Hu, 1991; Li, 2005)
- Japanese (Chew, 1969) Goes the other way!

Do listeners apply their own language's mapping when hearing speakers of other languages? Does it matter if it's the more or less common mapping? How can these perceptions illuminate or be illuminated by the production and acquisition patterns discussed in Tutorial 1?

Sibilants and gender

Several languages show gendering of /s/ and/or /ʃ/:

- English (Campbell-Kibler, 2011; Crist, 1997; Linville, 1998)
- Danish (Pharao and colleagues, in progress)
- Mandarin (Hu, 1991; Li, 2005)
- Japanese (Chew, 1969) Goes the other way!

Do listeners apply their own language's mapping when hearing speakers of other languages? Does it matter if it's the more or less common mapping? How can these perceptions illuminate or be illuminated by the production and acquisition patterns discussed in Tutorial 1?

This is a shared project across the Sound Size collective, with Mark Beckman, Kiwako Ito, Kathryn Campbell-Kibler, Tsz-Him Tsui, Eric Ruppe and Liz McCullough. Eric and I have been primarily working on this phase.

Problem 1: Data collection

- Open-end response
- Word-labelled ratings
- Numbered ratings
- Visual Analog Scale (VAS)
- Magnitude estimation

Problem 1: Data collection

- Open-end response
- Word-labelled ratings
- Numbered ratings
- Visual Analog Scale (VAS)
- Magnitude estimation

The first half of our problem is collecting appropriate data. Several tools have been used for this. Because we're discussing across multiple projects, we'll skip over the issues of question design that would distinguish, say, a semantic differential from a Likert scale. Our question here is the task given to the participant to conceptualize and convey their assessment.

Problem 1: Data collection

- Open-end response
- Word-labelled ratings
- Numbered ratings
- Visual Analog Scale (VAS)
- Magnitude estimation

- Open-end response
- Word-labelled ratings
- Numbered ratings
- Visual Analog Scale (VAS)
- Magnitude estimation

The first half of our problem is collecting appropriate data. Several tools have been used for this. Because we're discussing across multiple projects, we'll skip over the issues of question design that would distinguish, say, a semantic differential from a Likert scale. Our question here is the task given to the participant to conceptualize and convey their assessment.

Problem 1: Data collection

Visual Analog Scale (VAS)

- Easily analyzed
- Continuous responses
- Avoids number bias
- Simple to explain and perform

Visual Analog Scale (VAS)
- Easily analyzed
- Continuous responses
- Avoids number bias
- Simple to explain and perform

We have chosen the VAS for a few reasons. The continuous nature is particularly valuable because it enables analysis of patterns of individual responses, for example to ask whether a given listener is showing a unimodal or bimodal pattern, or to correlate responses from a single participant or two a single stimulus with other measures, e.g. IAT scores or formant measures. The “simple to explain and use” aspect has been field-tested by Heather Buchan, at the University of Wollongong who is working on acquisition of Gurundji Kriol and has used VAS with non-literate speakers with success.

Populations

- Subject pool (In person)
  - Closer control/observation
  - Free (if you have one)
  - Longer tasks possible
  - Wide range of tools possible
  - Limited population
- Mechanical Turk
  - Wider demographics
  - Low cost
  - Less control
  - Higher non-completion
  - Limited to online interface

Populations

- Subject pool (In person)
  - Closer control/observation
  - Free (if you have one)
  - Longer tasks possible
  - Wide range of tools possible
  - Limited population
- Mechanical Turk
  - Wider demographics
  - Low cost
  - Less control
  - Higher non-completion
  - Limited to online interface

We are collecting data in two venues: in the lab with a subject pool and on Mechanical Turk.

In-lab tools

- E-Prime — http://www.pstnet.com/eprime.cfm
- PsyScope — http://psy.ck.sissa.it/
- OpenSesame — http://www.cogsci.nl/software/opensesame

In-lab tools

- E-Prime — http://www.pstnet.com/eprime.cfm
- PsyScope — http://psy.ck.sissa.it/
- OpenSesame — http://www.cogsci.nl/software/opensesame

In person, we can use a wide range of tools, including paper and pencil tasks, face-to-face interviews, etc. For a VAS task, however, a computer interface is likely to our most reliable bet. There are a number of reliable suites out there. We've used E-Prime, but would like to hear others' preferences. There is a significant time and sometimes money cost associated with switching and E-Prime is a major choice at least in part because it's what we have! Here's a look at the tasks.

In-lab tools

- E-Prime — http://www.pstnet.com/eprime.cfm
- PsyScope — http://psy.ck.sissa.it/
- OpenSesame — http://www.cogsci.nl/software/opensesame

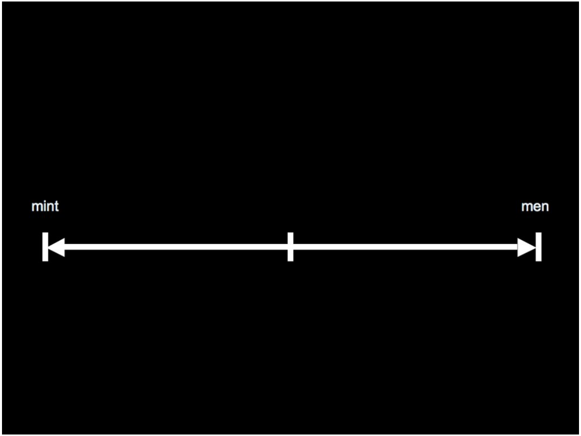
In-lab tools

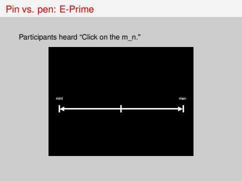
- E-Prime — http://www.pstnet.com/eprime.cfm
- PsyScope — http://psy.ck.sissa.it/
- OpenSesame — http://www.cogsci.nl/software/opensesame

In person, we can use a wide range of tools, including paper and pencil tasks, face-to-face interviews, etc. For a VAS task, however, a computer interface is likely to our most reliable bet. There are a number of reliable suites out there. We've used E-Prime, but would like to hear others' preferences. There is a significant time and sometimes money cost associated with switching and E-Prime is a major choice at least in part because it's what we have! Here's a look at the tasks.

Pin vs. pen: E-Prime

Participants heard “Click on the m_n."





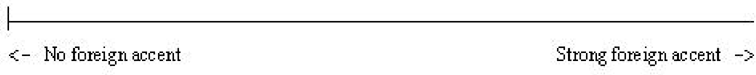
Here’s the slide we used for the pin vs. pen task.

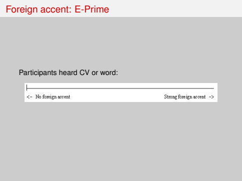
Face-voice connections

Face photos like this were rated for how accented, masculine, and educated they looked.

Foreign accent: E-Prime

Participants heard CV or word:

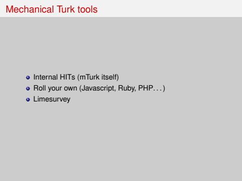




This paradigm uses clips of non-native English speech.

Mechanical Turk tools

- Internal HITs (mTurk itself)
- Roll your own (Javascript, Ruby, PHP. . .)
- Limesurvey



Depending on your needs and skills, there are different technical approaches to Mechanical Turk.

Mechanical Turk tools

- Internal HITs (mTurk itself)
- Roll your own (Javascript, Ruby, PHP. . .)
- Limesurvey

- Internal HITs (mTurk itself)
- Roll your own (Javascript, Ruby, PHP…)
- Limesurvey

Depending on your needs and skills, there are different technical approaches to Mechanical Turk.

LimeSurvey

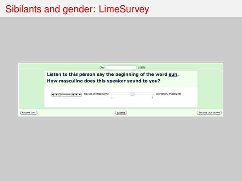
- Open-source PHP web application
- Easy to install
- Offers many different question types. For example
  - Open-text
  - Multiple-choice & radio buttons
  - Ranking
  - Sliders
- TurkLime links mTurk to LimeSurvey

LimeSurvey

- Open-source PHP web application
- Easy to install
- Offers many different question types. For example
  - Open-text
  - Multiple-choice & radio buttons
  - Ranking
  - Sliders
- TurkLime links mTurk to LimeSurvey

We're still arguing with LimeSurvey to get it to talk nicely to mTurk.

Sibilants and gender: LimeSurvey



So here's what our LimeSurvey question looks like. This appears within the Mechanical Turk frame:

Problem 2

Analyzing distributions which:

- happen on a bounded interval.
- are often multimodal
- $\rightarrow$ are not very normal!

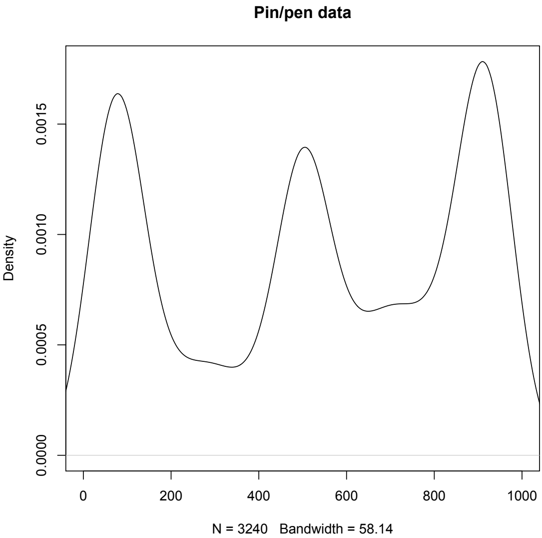
Problem 2

Analyzing distributions which:

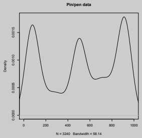
- happen on a bounded interval.
- are often multimodal
- → are not very normal!

One you've collected your data, you have a pile of data! It turns out that data of these types, when collected using VAS, are somewhat funky-shaped piles.

Pin/pen merger

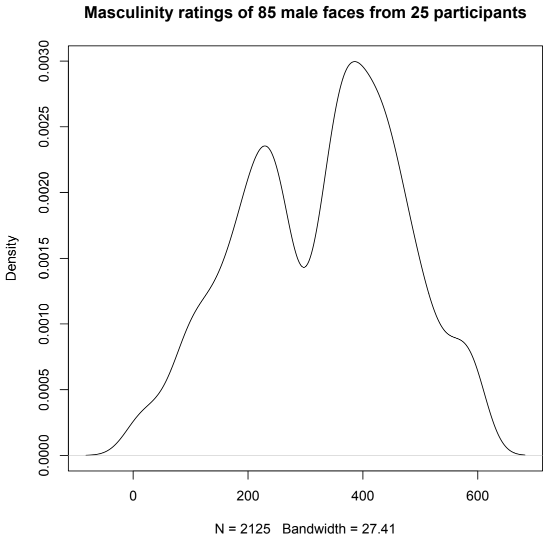


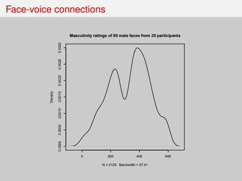
Pin/pen merger



The pin/pen data distribution. The three modes suggest that we did a good job of getting some merged and some non-merged speakers. But if we want to do some modeling to see, for example, how formant values predict the responses, it makes things challenging.

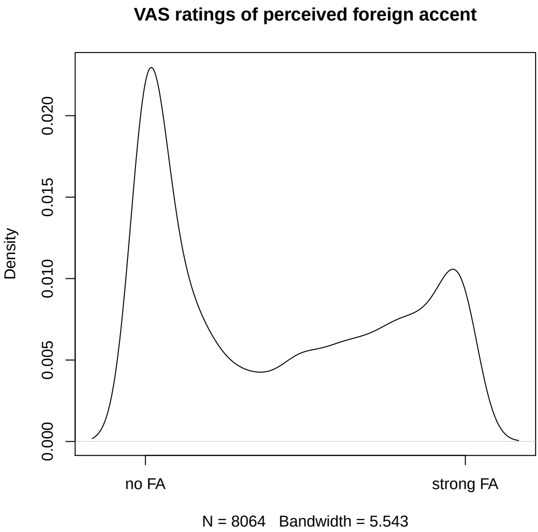
Face-voice connections



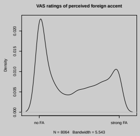


Two is a bit better than three, but still a challenge. Note also that the two modes are closer to the center than the outer modes for pin/pen.

Foreign accent



Foreign accent



And here.

Possible approaches

- Linear regression
- logit
- Beta regression
- Finite mixture models
- Beta mixture models

Possible approaches

- Linear regression
- logit
- Beta regression
- Finite mixture models
- Beta mixture models

So there are a few things we can do with these. One is that we can pretend that they are normally distributed even though they're not. This is really a stretch, given both the boundedness and the multi-modality. We can deal with the bounded issue by standardizing to 0-1 and taking the logit (log-odds) or by treating the distribution as a beta distribution. For multi-modal distributions, however, something else is needed, like a mixture model. Note that the type of distribution (e.g. normal, beta, etc) and number of underlying distributions are two different issues.

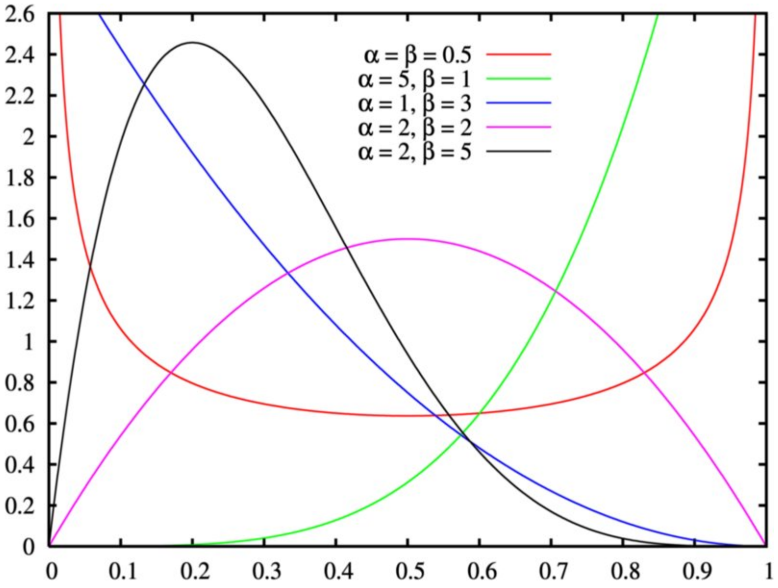
Possible approaches

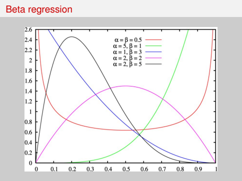
- Linear regression
- logit
- Beta regression
- Finite mixture models
- Beta mixture models

- Linear regression
- logit
- Beta regression
- Finite mixture models
- Beta mixture models

So there are a few things we can do with these. One is that we can pretend that they are normally distributed even though they're not. This is really a stretch, given both the boundedness and the multi-modality. We can deal with the bounded issue by standardizing to 0-1 and taking the logit (log-odds) or by treating the distribution as a beta distribution. For multi-modal distributions, however, something else is needed, like a mixture model. Note that the type of distribution (e.g. normal, beta, etc) and number of underlying distributions are two different issues.

Beta regression





Beta regression has been increasingly used in social science statistics, due to the prevalence of probability or probability-like distributions, e.g. poverty rates in a city. Beta distributions have two parameters (α and β) which do not straightforwardly map to real-world parameters of interest but variants on beta regression have been develop which estimate a mean and a dispersion parameter (). For certain types of VAS tasks, a beta distribution is likely to be very useful, given its range between a more central unimodal shape and a bimodal shape with the modes at the extremes.

Beta regression

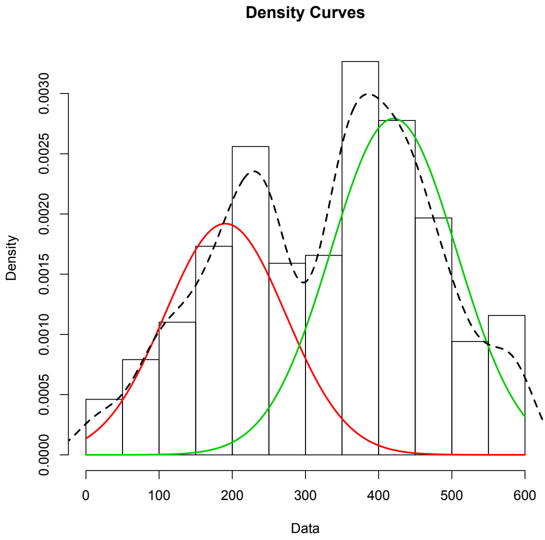
- Captures bounded range
- Captures potential unimodal or bimodal character?
- Intuitively plausible distribution
- Multiple modes must be at extremes

Beta regression

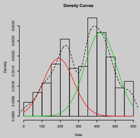
- Captures bounded range
- Captures potential unimodal or bimodal character?
- Intuitively plausible distribution
- Multiple modes must be at extremes

The R package *betareg* with perform beta regression to estimate mean and dispersion parameters. It's not able to manage mixed effects models (or mixture models), but there is code for BUGS (Bayesian inference Using Gibbs Sampling) which will do mixed-effects beta regression.

Finite Mixture Models



Finite Mixture Models



I know very little about finite mixture models, or their various cousins. But they seem like they might be a good choice here.

Finite Mixture Models

- Captures multi-modal character
- Intuitively plausible treatment: Participants are categorizing, then tinkering
- Characteristics depend on sub-distributions

Beta Mixture Models

- Best of both worlds?
- People in genomics like it!
- . . . ?

Our Questions

Where might micro-social judgments be most useful? Where does their utility stop?

- How can we maximize reliability and ecological validity given that we're only asking one question?
- What are the limits on how many times this task can be performed meaningfully?
-

Our Questions

How can researchers implement socio-friendly, experimentally sound studies online with a minimum of (previous) programming expertise?

- What would an ideal interface look like?
- How can we achieve it easily?
- Can existing tools be expanded to meet our needs?

Our Questions

What are the most effective tools for analyzing VAS data?

- How can we capture the multi-modality?
- Should we transform the data to manage the bounded distributions or is a beta distribution the right choice?
- How accessible are the possibly right tools (e.g. finite mixture models)?
- Are there ways we can help others understand and use them? (After we help ourselves do so!)

Brennan, Eileen M. and John S. Brennan. (1981). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research* 10(5):487–501.

Campanella, Salvatore and Pascal Belin. (2007). Integrating face and voice in person perception. *TRENDS in Cognitive Sciences* 11(12):535–543.

Campbell-Kibler, Kathryn. (2011). The sociolinguistic variant as a carrier of social meaning. *Language Variation and Change* 22(3):423–441.

Chew, Jr., John J.. (1969). The structure of japanese baby talk. *The Journal-Newsletter of the Association of Teachers of Japanese* 6(1):4–17.

Crist, Sean. (1997). Duration of onset consonants in gay male stereotyped speech. *University of Pennsylvania Working Papers in Linguistics* 4.3.

Dahan, Delphine, Sarah J. Drucker, and Rebecca A. Scarborough. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108:710–718.

de Gelder, Beatrice, Gilles Pourtois, and Lawrence Weiskrantz. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences of the USA* 99(6):4121–4126.

Gluszek, A. and J. F. Dovidio. (2010). The way they speak: A social psychological perspective on the stigma of non-native accents in communication. *Personality and Social Psychology Review* 14:215–237.

Hay, Jennifer, Aaron Nolan, and Katie Drager. (2006). From *fush* to *feesh*: Exemplar priming in speech perception. *The Linguistic Review* 23(3):351–379.

Hu, Minyang. (1991). Feminine accent in the Beijing vernacular. *Journal of the Chinese Language Teachers Association* 1:49–54.

Jacoby, Larry L.. (1991). A process dissociation framework:

Separating automatic from intentional uses of memory. *Journal of Memory and Language* 30(5):513–541.

Kinzler, Katherine D., Emmanuel Dupoux, and Elizabeth S. Spelke. (2007). The native language of social cognition. *PNAS* 104(30):12577–12580.

Li, Fangfang. (2005). An acoustic study on feminine accent in the Beijing dialect. In Q. Gao (Ed.), *Proceedings of the 17th North American Conference on Chinese Linguistics*. Los Angeles: (Graduate Students in Linguistics), University of Southern California.

Linville, Sue Ellen. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia Phoniatrica et Logopaedica* 50(1):35–48.

Payne, B. Keith. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology* 89(4):488–503.

Payne, B. Keith and Brandon D. Stewart. (2007). Automatic and controlled components of social cognition: A process dissociation approach. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*. New York: Psychology Press.

Purnell, Thomas, William Idsardi, and John Baugh. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18(1):10–30.

Staum Casasanto, Laura. (2008). Does social information influence sentence processing? In *30th Annual Meeting of the Cognitive Science Society* Washington, D.C.

Williams, Frederick, Nancy Hewett, Robert Hopper, Leslie M. Miller, Rita C. Naremore, and Jack L. Whitehead. (1976). *Explorations of the Linguistic Attitudes of Teachers*. Rowley, Mass.: Newbury House.