Nonparametric methods

Roger Levy

5 May 2025

Nonparametric methods

- You have probably encountered the term "nonparametric" (statistics/models/methods/...), and likely will again
- There are **two** rather different meanings that "nonparametric" means in the context of statistics:
- Methods that **do not assume a distributional form** underlying the data
  - E.g., sign test; Wilcoxon signed-rank test; the bootstrap
- Methods that involve parameters, but whose **expressive flexibility grows with data scale and complexity**
  - E.g., kernel density estimation; *k* nearest neighbors; certain hierarchical Bayesian models; GAMs
- Wikipedia's nice "parent" characterization:

Nonparametric statistics

文A 24 languages

From Wikipedia, the free encyclopedia

Nonparametric statistics is a type of statistical analysis that makes minimal assumptions about the underlying distribution of the data being studied. Often these models are infinite-dimensional, rather than finite dimensional, as in parametric statistics.[1]

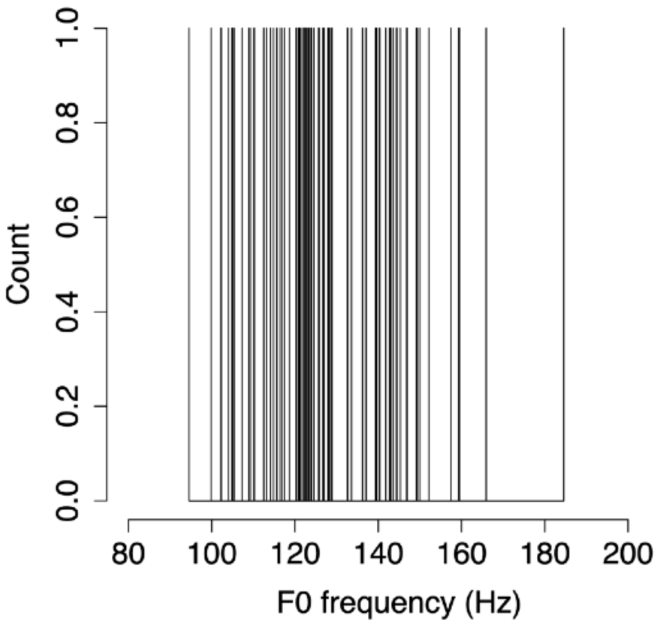
Example 1: kernel density estimation

Example 1: kernel density estimation

- **Task:** estimate the distribution of adult male American English speakers' F0 ("pitch") formant for the vowel ɑ

Example 1: kernel density estimation

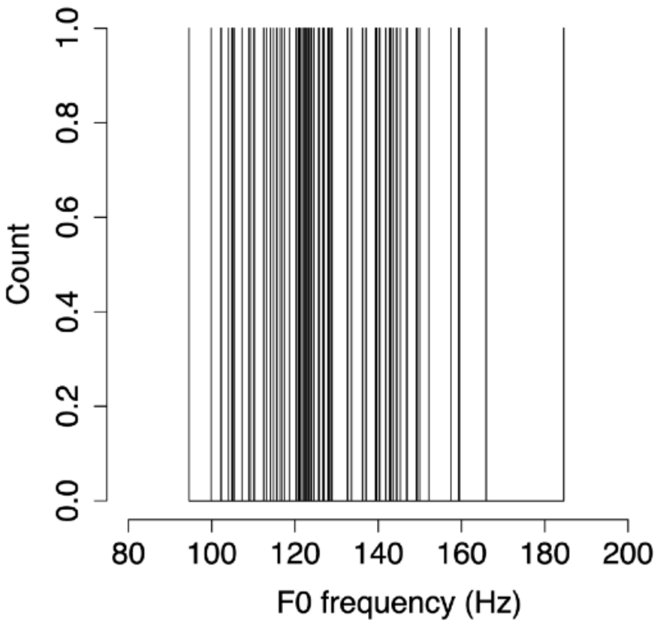
- **Task:** estimate the distribution of adult male American English speakers' F0 ("pitch") formant for the vowel ɑ



Raw data

Example 1: kernel density estimation

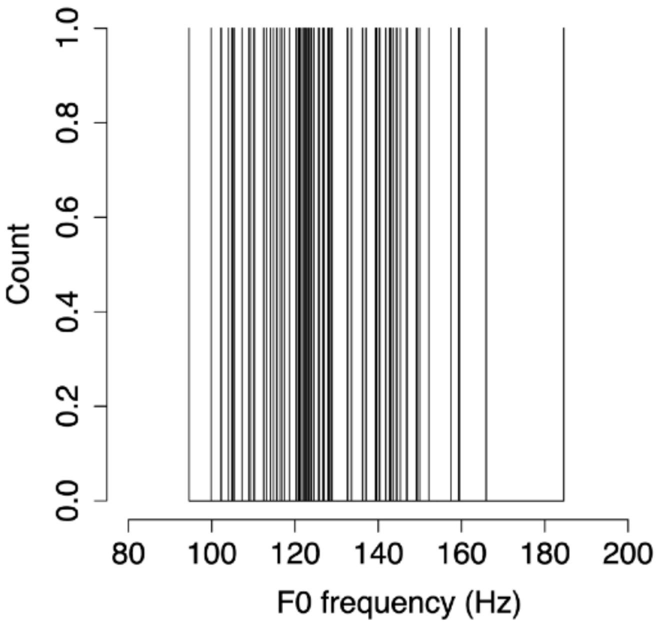
- **Task:** estimate the distribution of adult male American English speakers' F0 ("pitch") formant for the vowel ɑ
- With histograms:



Raw data

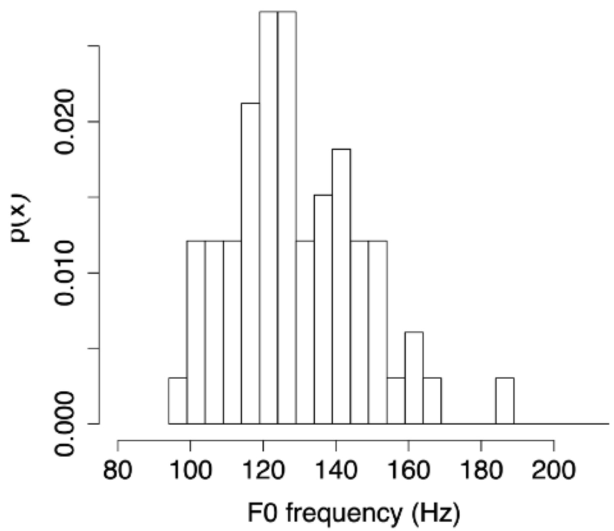
Example 1: kernel density estimation

- **Task:** estimate the distribution of adult male American English speakers' F0 ("pitch") formant for the vowel ɑ
- With histograms:



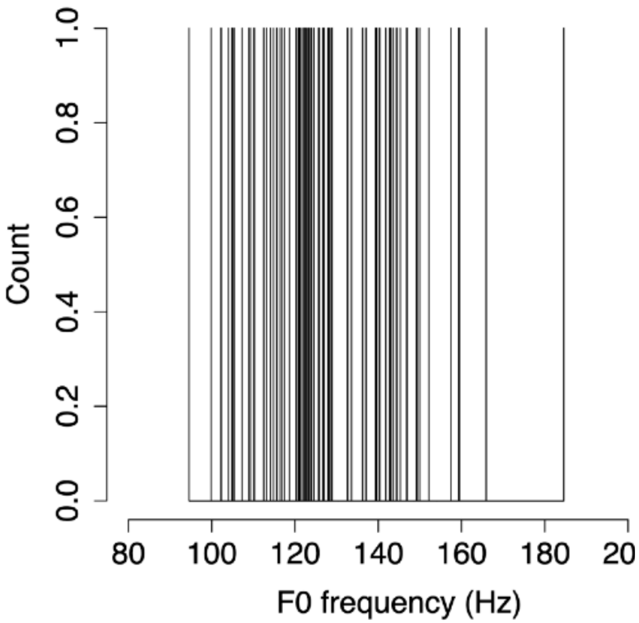
Raw data

5Hz-bin histogram starting at 95Hz

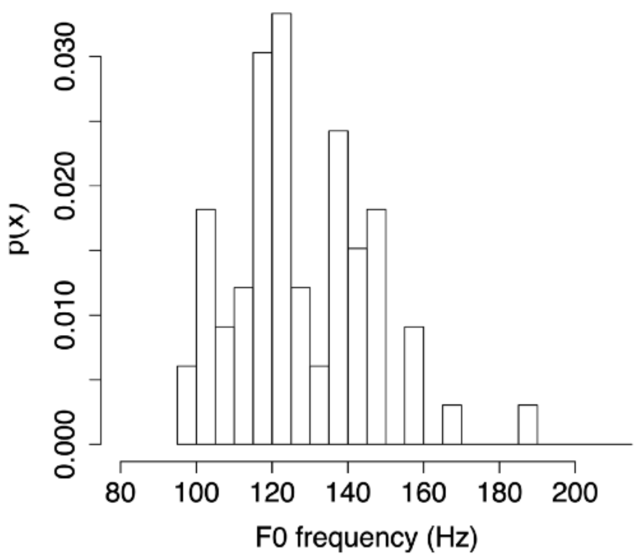


Example 1: kernel density estimation

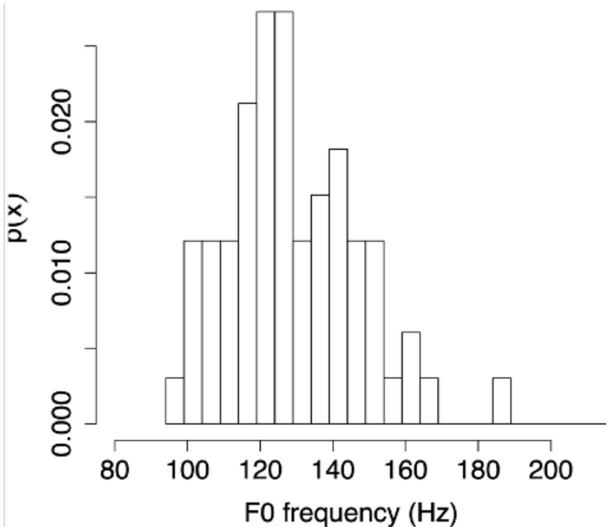
- **Task:** estimate the distribution of adult male American English speakers' F0 ("pitch") formant for the vowel ɑ
- With histograms:



Raw data



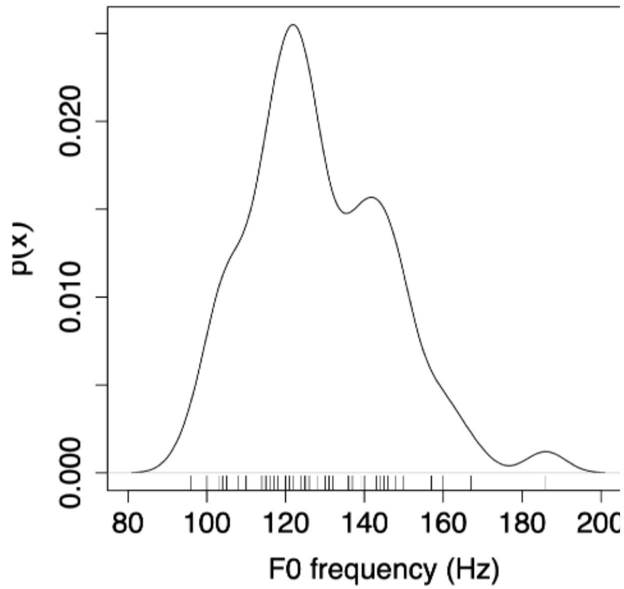
5Hz-bin histogram starting at 95Hz

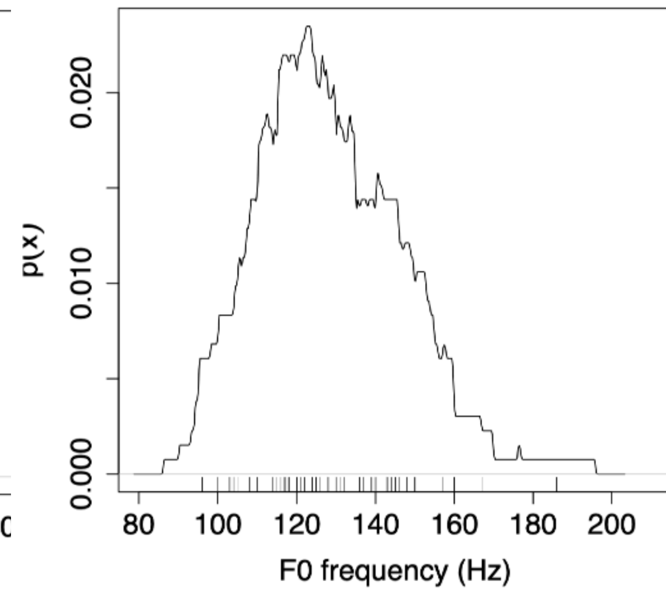


5Hz-bin histogram starting at 94Hz

Example 1: kernel density estimation

$$\hat{p}(X = x) = \frac{1}{n\sqrt{2\pi b^2}} \sum_{i=1}^{n} \exp\left[-\frac{(x - x_i)^2}{2b^2}\right]$$

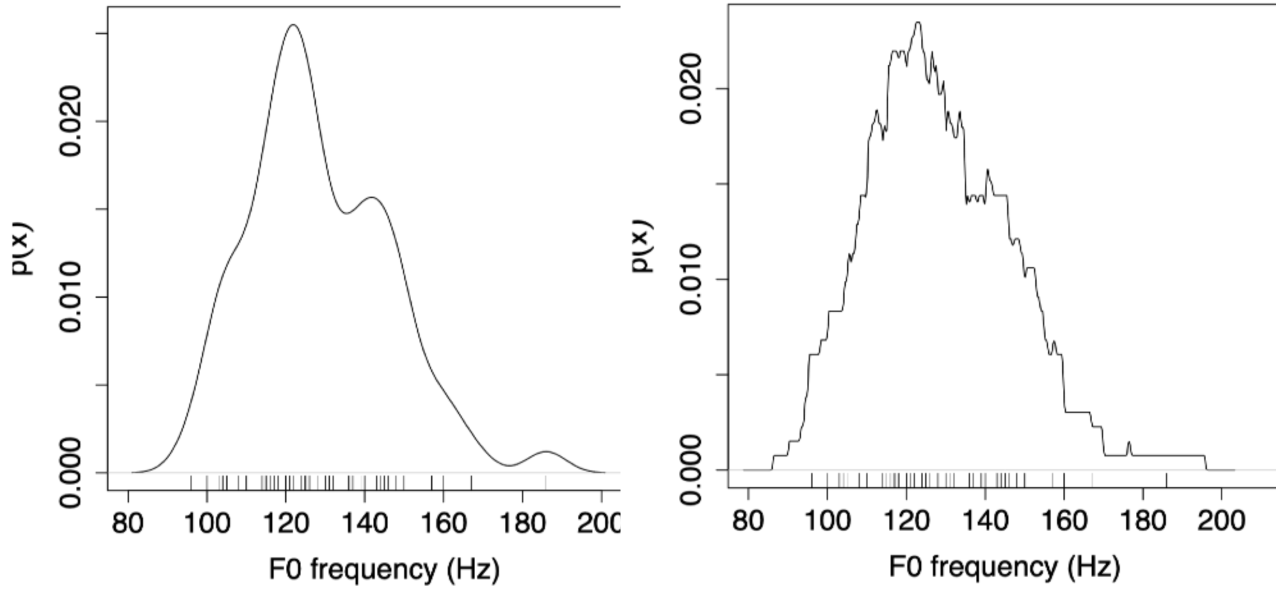




Example 1: kernel density estimation

- **Kernel:** an integrable non-negative real function K, typically **normalized** (integrates to 1) and **symmetric**
  - Examples: the **normal** and **rectangular** kernels
  - Normal kernel density estimate for **bandwidth** b:

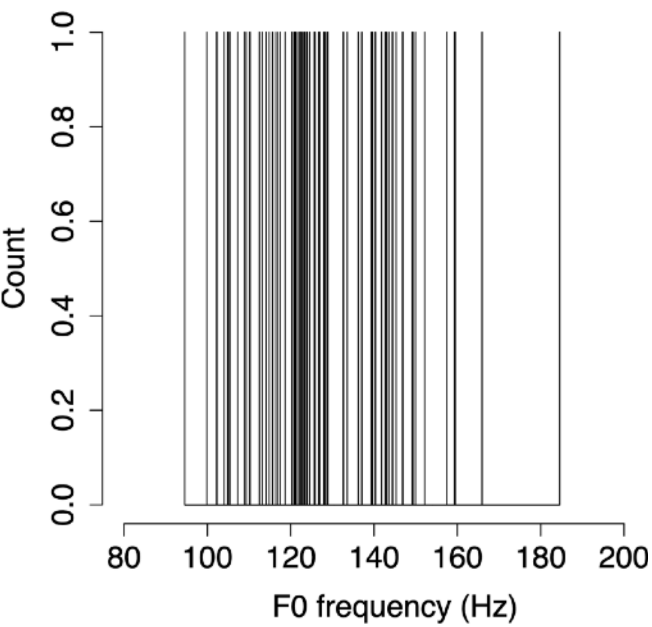
$$\hat{p}(X = x) = \frac{1}{n\sqrt{2\pi b^2}} \sum_{i=1}^{n} \exp\left[-\frac{(x - x_i)^2}{2b^2}\right]$$



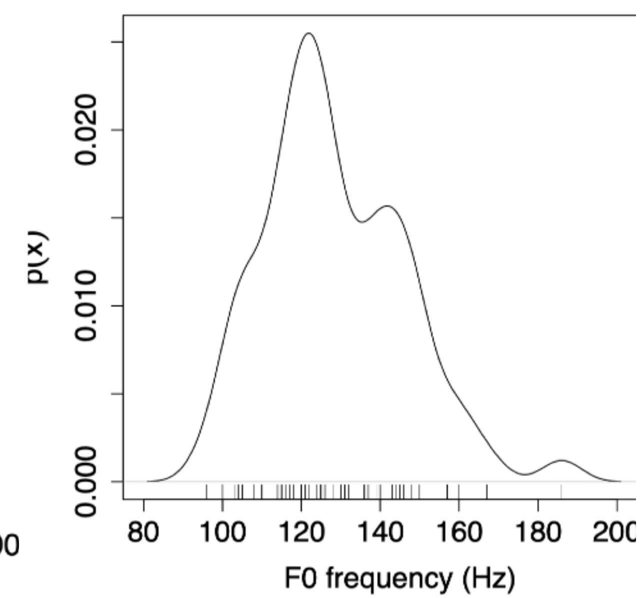
Example 1: kernel density estimation

- **Kernel:** an integrable non-negative real function K, typically **normalized** (integrates to 1) and **symmetric**
  - Examples: the **normal** and **rectangular** kernels
  - Normal kernel density estimate for **bandwidth** b:

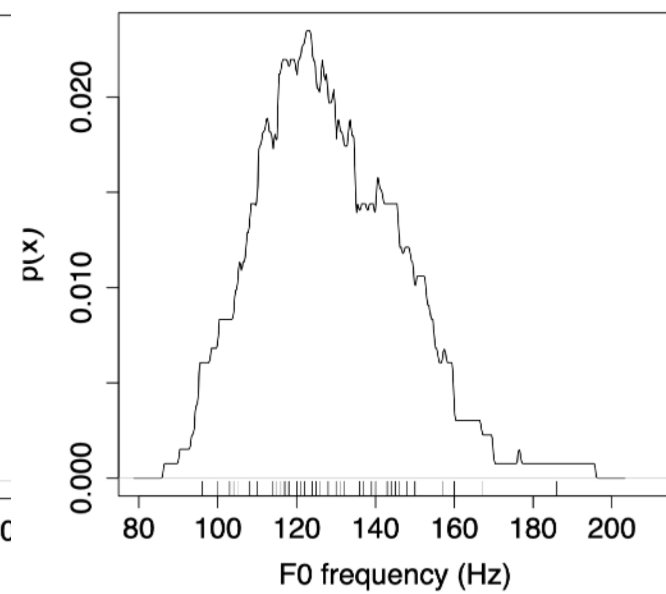
$$\hat{p}(X = x) = \frac{1}{n\sqrt{2\pi b^2}} \sum_{i=1}^{n} \exp\left[-\frac{(x - x_i)^2}{2b^2}\right]$$



Raw data



Normal kernel, $b = 5$



Rectangular kernel, $b = 20$

Example 2: the **sign test**

- Consider a vector of real-valued observations $y_1, \ldots y_n$
- Assume they are generated IID
- Null hypothesis H_0: $P(y > 0) = 0.5$
- This is a special case of the **binomial test**

Example 3: Wilcoxon rank sum test

- Consider two real-valued samples, $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$, each IID
- Null hypothesis H_0: the two samples come from **the same distribution**

Connecting to parametric statistics

Connecting to parametric statistics

- What is the relationship between the t test, the sign test, and the rank sum test (also the signed rank test which we didn't cover just now)?

Connecting to parametric statistics

- What is the relationship between the t test, the sign test, and the rank sum test (also the signed rank test which we didn't cover just now)?
- **Mini-practicum:** When would you want to use one versus the other? What are the pitfalls involved?

The bootstrap

- Common use case: confidence intervals
- Previously in this class, we have used **parametric assumptions** (typically some kind of normality) to compute confidence intervals
- When these assumptions are wrong, it can affect our confidence intervals
- The idea of the bootstrap: **use the distributional properties of the data itself to estimate the statistic you're interested in**
- Simplest use may be **case resampling**
- **Mini-practicum:** implement case resampling for estimating confidence intervals on non-normally (and normally, for sanity checking) distributed data