Zipf's power law in activity schedules and the effect of aggregation

Wim Ectors[a] Bruno Kochan[a] Davy Janssens[a] Tom Bellemans[a]
Geert Wets[a]

[a]Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium

18-05-2017

ANT 2017

Overview

Overview

IMOB TRANSPORTATION RESEARCH INSTITUTE
UHASSELT

Intro: Zipf's law?

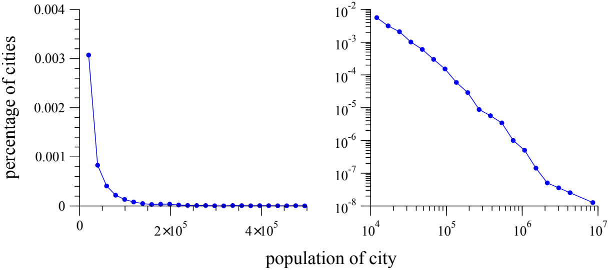
Complex observations, but simple empirical distribution

- Power-law distribution

$$p(x) = Cx^{-\alpha}$$

- Rank-size interpretation

$$f(r_i) = \frac{f(r_1)}{r_i}$$

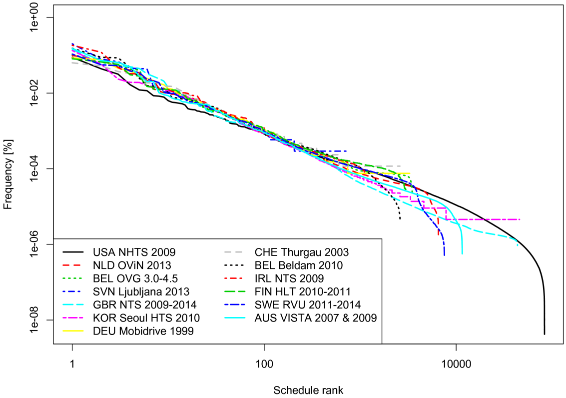


(Newman, 2005)

- Observed in many fields (Newman, 2005; Okuyamaet al., 1999; Zipf, 1949)
    - Word frequency
    - City size
    - Earthquake magnitude
    - Annual company income
    - Solar flares
    - Number of citations of papers
    - ...
- Traffic demand models
    - Extremely complex, multidimensional processes
    - AB-models
        - (non-)mandatory activity schedulers
        - TOD models
        - re-schedulers
        - interaction with others
        - ...
- *Yet, day-long schedules obey a simple power law*

In a previous research:

- Universal power law governs schedule data (13 HTSs from across the world)



This research:

- Aggregation of Zipf's law in activity schedules?
  - Dimension 1: activity type encoding
  - Dimension 2: aggregation of individual data

IMOB TRANSPORTATION RESEARCH INSTITUTE UHASSELT

Overview

1. Intro: Zipf's law?
2. Data & Estimation Procedure
3. Aggregation of Activity Type Encoding
4. Aggregation of Individual Data
5. Conclusion

Data & Estimation Procedure

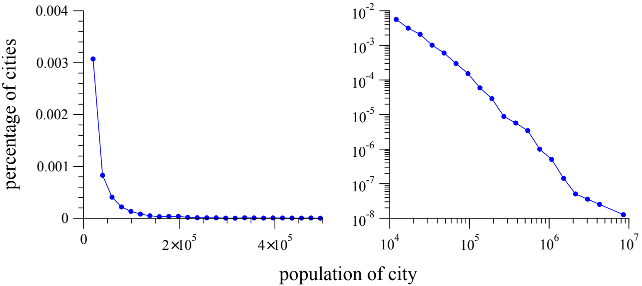
4 data sets:

- USA NHTS 2009 (1-day HTS; 258k schedules)
- DEU MobiDrive 1999 (6-week HTS; 13k schedules)
- CHE Thurgau 2003 (6-week HTS; 9k schedules)
- Donated trip data of one individual (24 weeks)
  - Moves app; enriched with trip purposes (10 classes)

Out-of-home activity schedules created by concatenating trip purposes into a sequence

- *"home-work-shopping-home-sport-home"*

Data & Estimation Procedure



(Newman, 2005)

Fitting a power law?

- X linear regression (least squares) to log-transformed variables
- ✓ MLE with Kolmogorov-Smirnov (KS) gof as cutoff criterion
- *R* package PoweRlaw (Gillespie, 2015)
  - bootstrapping procedure evaluates parameter estimation uncertainty

Overview

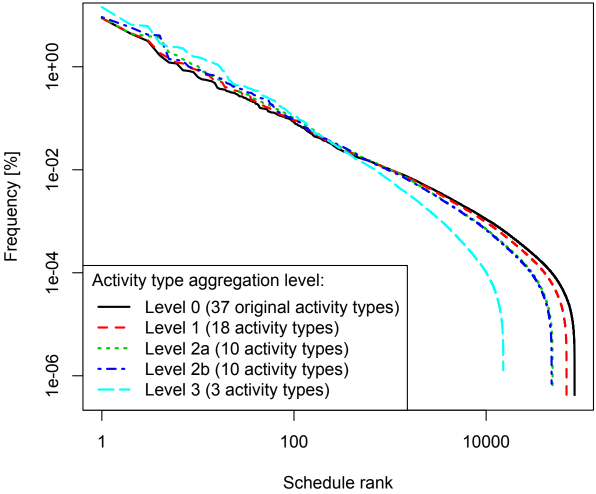
1. Intro: Zipf's law?
2. Data & Estimation Procedure
3. Aggregation of Activity Type Encoding
4. Aggregation of Individual Data
5. Conclusion

Aggregation of Activity Type Encoding

- USA NHTS 2009 data
- Originally 37 distinct trip purposes
  Aggregated to:
  - 18 act. type
  - 10 act. types (2x)
  - 3 act. types
- Still power-law distributed? (with the same exponent?)

TRANSPORTATION RESEARCH INSTITUTE
UHASSELT

Aggregation of Activity Type Encoding



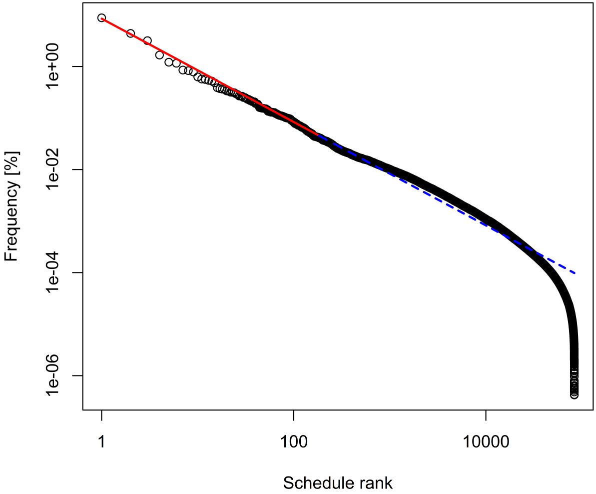
	poweRlaw estimations (MLE + KS)			Bootstrapping uncertainty evaluation		
Aggregation or subset	α	x_{min}	Cum. pct rejected	AM(α)	SD(α)	P-value
Level 0 (37 original activity types)	2.003	36809977	55%	2.006	0.070	0.255
Level 1 (18 activity types)	1.967	36837451	50%	1.972	0.065	0.166
Level 2a (10 activity types)	1.934	46135634	43%	1.939	0.065	0.998
Level 2b (10 activity types)	1.892	60781076	45%	1.899	0.071	0.741
Level 3 (3 activity types)	1.890	109512566	28%	1.891	0.084	0.835

Data: USA NHTS 2009.

Overview

1. Intro: Zipf's law?
2. Data & Estimation Procedure
3. Aggregation of Activity Type Encoding
4. Aggregation of Individual Data
   - Study Area Level
   - Subsets of Study Area
   - The Individual Level
5. Conclusion

Study Area Level



Aggregation of Individual Data

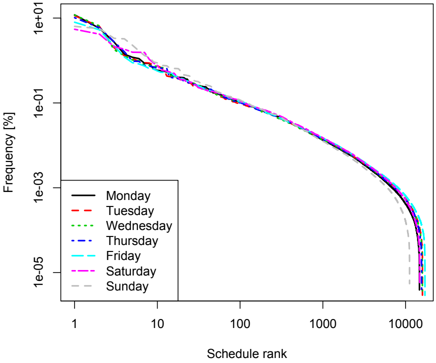
	poweRlaw estimations (MLE + KS)			Bootstrapping uncertainty evaluation		
Dataset	α	x_{min}	Cum. pct rejected	AM(α)	SD(α)	P-value
USA NHTS 2009	2.003	36809977	55%	2.006	0.070	0.255
DEU Mobidrive 1999	2.053	23	52%	2.002	0.133	0.714
CHE Thurgau 2003	1.929	16	49%	2.009	0.113	0.317

Note: the different scales of x_{min} are caused by different weight variables.

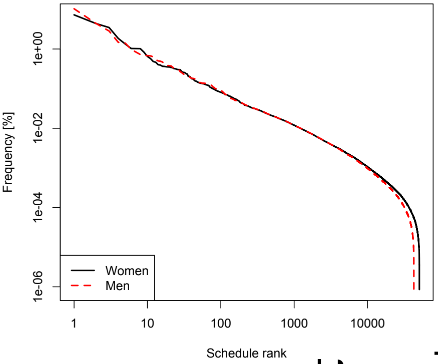
Subsets of Study Area

- USA NHTS 2009 data
  - sufficiently large
- Subsets based on
  - DOW
  - Gender

Subsets of Study Area



Day of the week



Gender

Aggregation of Individual Data

	poweRlaw estimations (MLE + KS)			Bootstrapping uncertainty evaluation		
Aggregation or subset	α	x_{min}	Cum. pct rejected	AM(α)	SD(α)	P-value
All individuals aggregated	2.003	36809977	55%	2.006	0.070	0.255
Monday*	2.290	46616705	67%	2.270	0.359	0.831
Tuesday*	2.161	35581917	67%	2.182	0.236	0.820
Wednesday*	2.152	45646004	68%	2.172	0.267	0.679
Thursday*	2.088	48120314	71%	2.140	0.282	0.221
Friday*	2.279	34509610	72%	2.284	0.250	0.901
Saturday*	2.182	61045896	76%	2.176	0.288	0.134
Sunday*	2.091	52160661	66%	2.060	0.200	0.982
Women	2.104	37421218	61%	2.115	0.114	0.551
Men	2.157	36416801	58%	2.165	0.116	0.783

Data: USA NHTS 2009.

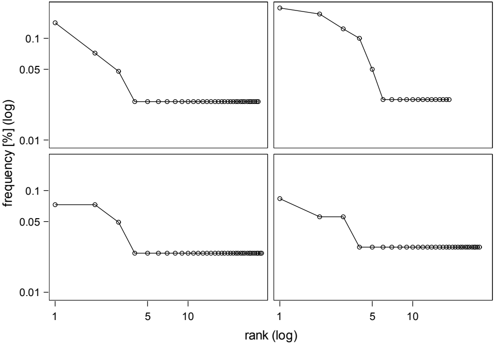
The Individual Level

- Challenging...
- DEU MobiDrive 1999 (6 weeks)
- CHE Thurgau 2003 (6 weeks)
- Donated trip data of one individual (24 weeks)

Do the schedules of each individual exhibit a power law distribution?

- KS goodness-of-fit test

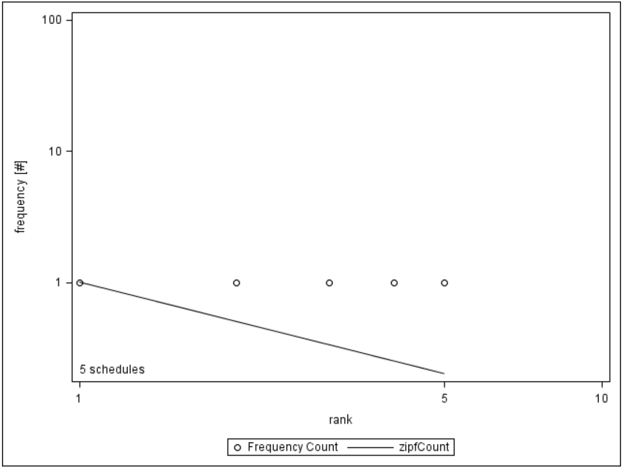
The Individual Level



Examples of distributions rejecting H_0: power law is a good fit

- MobiDrive & Thurgau
- Only 12% do not reject H_0 of a power law distribution
- $\rightarrow$ results do **not** support Zipf's law
- However, *insufficient* data?

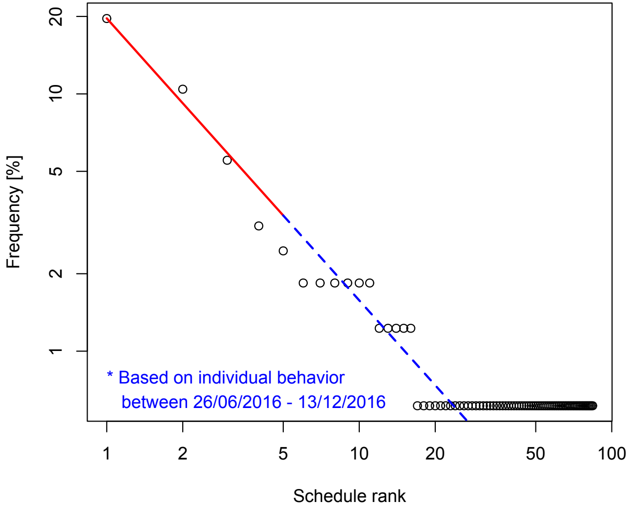
The Individual Level



The Individual Level

- $n \gtrsim 50$ needed, but $\overline{n} = 37.625 < 50$ (Clauset et al., 2009)
- Those rejecting H_0: transition phase
- Those not rejecting H_0: already quite advanced evolution
- More than 6 weeks may be needed
  - Donated trip data of one individual (24 weeks)

The Individual Level



Aggregation of Individual Data

	poweRlaw estimations (MLE + KS)			Bootstrapping uncertainty evaluation		
Dataset	α	$x_{\min}$	Cum. pct rejected	AM(α)	SD(α)	P-value
Donated schedules from an individual	2.454	4	59%	2.629	0.708	0.689

Overview

Conclusion

- People's behavior very complex, but schedules power law distributed
- No significant influence of activity type encoding aggregation
  - Only at extreme levels of aggregation may the power law break down more quickly
- Study area → subsets → individual data
  - Power law is found quite consistently ...
  - On the condition that sufficient individual data is available

Further research:

- Correlate stage of evolution of power law to person characteristics
- Modeling the mechanism leading to Zipf's law

References

- Newman, M.: Power Laws, Pareto Distributions and Zipf's Law. Contemporary physics. 46, 323-351 (2005). doi:10.1080/00107510500052444
- Okuyama, K., Takayasu, M., Takayasu, H.: Zipf's law in income distribution of companies. Physica A: Statistical Mechanics and its Applications. 269, 125-131 (1999). doi:10.1016/S0378-4371(99)00086-2
- Zipf, G.K.: Human Behaviour and the Principle of Least Effort. Addison-Wesley, Reading (1949)
- Gillespie, C.S.: Fitting Heavy Tailed Distributions: The poweRlaw Package. Journal of Statistical Software. 64, 1-16 (2015)
- Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-Law Distributions in Empirical Data. SIAM Review. 51, 661 (2009). doi:10.1137/070710111
- U.S. Department of Transportation, Federal Highway Administration: 2009 National Household Travel Survey, http://nhts.ornl.gov, (2009)
- Chalasani, V.S., Axhausen, K.W.: Mobidrive: A six week travel diary, https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/tsms/tsms2.pdf, (2004)
- Loechl, M.: Stability of Travel Behaviour: Thurgau 2003, http://archiv.ivt.ethz.ch/vpl/publications/tsms/tsms16.pdf, (2005)

Thank you!