Zipf’s law in activity schedules

Wim Ectors, Bruno Kochan, Davy Janssens, Tom Bellemans & Geert Wets

hEART 2016
Table of contents

- Introduction
  - Zipf’s law?
- Methodology
- Aims
- Results
- Discussion and conclusion
Introduction

- 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 distribution: Zipf’s law \((Zipf, 1949)\)
Introduction – Zipf’s law

- Power-law distribution
  - Zipf: $\alpha=1$
    
    $$p(x) = Cx^{-\alpha}$$

- Rank-size interpretation
  
  $$f(r_i) = \frac{f(r_1)}{r_i}$$

Introduction – Zipf’s law

- Observed in many different fields:
  - Word frequency
  - City size
  - Earthquake magnitude
  - Annual company income
  - Solar flares
  - Number of citations of papers
  - Fujiwara, 2004; Furusawa and Kaneko, 2003; Maillart et al., 2008; Newman, 2005; Okuyama et al., 1999

- General applicability still sometimes contested:
  - City size (Soo, 2005)
Introduction – Zipf’s law

- In domain of transportation
  - Power-law like distributions in
    - Displacement distance
    - Gyration radius
    - Location visiting frequency
    - Location visiting duration
    - Bus transport networks

- This study:
  - Day-long activity pattern (=schedule) frequency
Methodology

- Zipf’s law in this study:
  - Schedule frequency
    1. Compose schedules from HTS data
    2. One-way frequency table
    3. Sort & add rank column
    4. Plot rank vs schedule frequency

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Freq. [%]</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-11-1</td>
<td>8.82</td>
<td>1</td>
</tr>
<tr>
<td>1-21-1</td>
<td>4.39</td>
<td>2</td>
</tr>
<tr>
<td>1-41-1</td>
<td>3.17</td>
<td>3</td>
</tr>
<tr>
<td>1-53-1</td>
<td>1.67</td>
<td>4</td>
</tr>
<tr>
<td>1-51-1</td>
<td>1.21</td>
<td>5</td>
</tr>
<tr>
<td>1-22-1</td>
<td>1.16</td>
<td>6</td>
</tr>
<tr>
<td>1-30-1</td>
<td>0.85</td>
<td>7</td>
</tr>
<tr>
<td>1-82-1</td>
<td>0.83</td>
<td>8</td>
</tr>
<tr>
<td>1-41-41-1</td>
<td>0.78</td>
<td>9</td>
</tr>
<tr>
<td>1-54-1</td>
<td>0.63</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1: Top-10 schedules NHTS ‘09
Methodology

Observed frequencies and ML + KS power-law fit in NHTS 2009 schedules
Methodology

- ML fit
- Kolmogorov-Smirnov cutoff criterion
  - Power-law for values $> x_{min}$
- $R$ package: poweRlaw  
  (Gillespie 2015)
Aims

1. To provide evidence for a rank-frequency Zipf’s law in activity schedule frequencies
2. To test the law’s dependency on the aggregation level of activity types
3. To test the law’s validity for each day of the week
1. Providing evidence for a rank-frequency Zipf’s law in activity schedule frequencies

<table>
<thead>
<tr>
<th>Dataset</th>
<th>poweRlaw estimations (ML + Kolmogorov-Smirnov)</th>
<th>Bootstrapping uncertainty evaluation (5000 (*2000) simulations)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α</td>
<td>xmin</td>
</tr>
<tr>
<td>NHTS 2009*</td>
<td>2.003</td>
<td>36809977</td>
</tr>
<tr>
<td>OViN 2013</td>
<td>1.885</td>
<td>1325</td>
</tr>
<tr>
<td>OViN 2013 (OVG encoding)</td>
<td>1.830</td>
<td>2378</td>
</tr>
<tr>
<td>OVG 3.0-4.5</td>
<td>1.947</td>
<td>2</td>
</tr>
<tr>
<td>Ljubljana (Slovenia)</td>
<td>1.995</td>
<td>9</td>
</tr>
<tr>
<td>UK nts 2009-2014*</td>
<td>1.862</td>
<td>4,071</td>
</tr>
<tr>
<td>Seoul MA 2010*</td>
<td>1.859</td>
<td>26</td>
</tr>
</tbody>
</table>
1. Providing evidence for a rank-frequency Zipf’s law in activity schedule frequencies
2. Testing the law’s dependency on the aggregation level of activity types

<table>
<thead>
<tr>
<th>Dataset</th>
<th># levels [encoding]</th>
<th>poweRlaw estimations (ML + Kolmogorov-Smirnov)</th>
<th>Bootstrapping uncertainty evaluation (5000 (*2000) simulations)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\alpha$</td>
<td>Xmin</td>
</tr>
<tr>
<td>NHTS 2009*</td>
<td>37 [“Original”]</td>
<td>2.003</td>
<td>36809977</td>
</tr>
<tr>
<td></td>
<td>18 [“Intermediate”]</td>
<td>1.967</td>
<td>36837451</td>
</tr>
<tr>
<td></td>
<td>10 [“OVG”]</td>
<td>1.934</td>
<td>46135634</td>
</tr>
<tr>
<td></td>
<td>10 [“Group”]</td>
<td>1.892</td>
<td>60781076</td>
</tr>
<tr>
<td></td>
<td>3 [“Cat”]</td>
<td>1.890</td>
<td>109512566</td>
</tr>
</tbody>
</table>
2. Testing the law’s dependency on the aggregation level of activity types
Results

3. Testing the law’s validity for each day of the week

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Subset</th>
<th>(\alpha)</th>
<th>(X_{\text{min}})</th>
<th>Cum. Pct discarded</th>
<th>(n_{\text{tail}})</th>
<th>AM</th>
<th>SD</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHTS 2009*</td>
<td>All data</td>
<td>2.003</td>
<td>36809977</td>
<td>55%</td>
<td>181</td>
<td>2.006</td>
<td>0.070</td>
<td>0.255</td>
</tr>
<tr>
<td></td>
<td>Monday</td>
<td>2.290</td>
<td>46616705</td>
<td>67%</td>
<td>22</td>
<td>2.270</td>
<td>0.359</td>
<td>0.831</td>
</tr>
<tr>
<td></td>
<td>Tuesday</td>
<td>2.161</td>
<td>35581917</td>
<td>67%</td>
<td>26</td>
<td>2.182</td>
<td>0.236</td>
<td>0.820</td>
</tr>
<tr>
<td></td>
<td>Wednesday</td>
<td>2.152</td>
<td>45646004</td>
<td>68%</td>
<td>20</td>
<td>2.172</td>
<td>0.267</td>
<td>0.679</td>
</tr>
<tr>
<td></td>
<td>Thursday</td>
<td>2.088</td>
<td>48120314</td>
<td>71%</td>
<td>17</td>
<td>2.140</td>
<td>0.282</td>
<td>0.221</td>
</tr>
<tr>
<td></td>
<td>Friday</td>
<td>2.279</td>
<td>34509610</td>
<td>72%</td>
<td>28</td>
<td>2.284</td>
<td>0.250</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>Saturday</td>
<td>2.182</td>
<td>61045896</td>
<td>76%</td>
<td>15</td>
<td>2.176</td>
<td>0.288</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td>Sunday</td>
<td>2.091</td>
<td>52160661</td>
<td>66%</td>
<td>21</td>
<td>2.060</td>
<td>0.200</td>
<td>0.982</td>
</tr>
</tbody>
</table>

Bootstrapping uncertainty evaluation (5000 (*2000) simulations)
3. Testing the law's validity for each day of the week
Discussion and conclusion

- A power-law distribution (Zipf) appears to be
  - valid for most schedules and to be study area independent
  - largely independent of activity type encoding aggregation
  - valid on different days of the week

- Further research on
  - Evidence general applicability (≠ study areas)
  - Explore the extent of validity (aggregated → disaggregated data)
  - Investigate origin
References

Thank you

Wim Ectors*
Bruno Kochan
Davy Janssens
Tom Bellemans
Geert Wets

*wim.ectors@uhasselt.be