Bayesian Digit Analysis

Pedro Fonseca

ISEG Lisbon School of Economics and Management, University of Lisbon
CEMAPRE – Center for Applied Mathematics and Economics

Background and Motivation

Digit analysis consists in using empirical regularities regarding the occurrence of digits in numbers to screen numerical datasets for erroneous or fraudulent data. Most of the digit analysis literature relies on Benford's law, a logarithmically-decaying pattern in leading digits frequencies that was first discovered by Newcomb (1881) and later supported with empirical evidence by Benford (1938). The first rigorous proof of the emergence of Benford's law is due to Hill (1995). According to Benford's law:

$$P(\text{First Digit} = d_1) = \log\left(1 + \tfrac{1}{d_1}\right),\ d_1 \in \{1, \dots, 9\} \qquad \text{(BL1)}$$

$$P(\text{Second Digit} = d_2) = \textstyle\sum_{d_1=1}^{9} \log\left(1 + \tfrac{1}{10\, d_1 + d_2}\right),\ d_2 \in \{0, \dots, 9\} \qquad \text{(BL2)}$$

Digit analysis usually requires testing point null hypotheses. In that context, classical significance tests are known to over-reject the null in large samples due to the high levels of power they attain, as the acceptance region shrinks with sample size for a given significance level (Ley, 1996). Consequently, a large proportion of legit datasets is likely to be classified as fraudulent. Any deviation from the null, no matter how tiny, can produce small p-values and be declared statistically significant if the sample is large enough (Wasserstein and Lazar, 2016). Hence, the usefulness and interpretation of classical p-values are drastically affected by sample size (Pericchi and Torres, 2011).

Methodology Part I

- Bayesian model selection: Multinomial ∧ Dirichlet model
  - Likelihood: $\boldsymbol{x}|\boldsymbol{\theta} \sim \mathcal{M}_k(N, \boldsymbol{\theta})$
  - Hypotheses: $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta_0}$ vs $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta_0}$.
  - Prior: $\pi(\boldsymbol{\theta}|H_0) = 1_{\boldsymbol{\theta_0}}(\boldsymbol{\theta_0})$ and $\boldsymbol{\theta}|H_1 \sim \text{Dir}_k(\boldsymbol{\alpha})$
    - (i) $\boldsymbol{\alpha} = \mathbf{1}$: Uniform prior
    - (ii) $\boldsymbol{\alpha} = c\,\boldsymbol{\theta_0}$, $c > 0$: Dirichlet prior centred on $\boldsymbol{\theta_0}$
  - Marginal density: $m_i(\boldsymbol{x}) = \int_{\Theta_i} f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|H_i)\, d\boldsymbol{\theta},\ i = 0, 1$
    - (i) $m_0(\boldsymbol{x}) = \frac{N!}{\prod_{i=1}^{k+1} x_i!} \prod_{i=1}^{k+1} {\theta_{0i}}^{x_i}$ (ii) $m_1(\boldsymbol{x}) = \frac{N!}{\prod_{i=1}^{k+1} x_i!} \frac{B(\boldsymbol{\alpha}+\boldsymbol{x})}{B(\boldsymbol{\alpha})}$
  - Bayes factor: $B_{01}(\boldsymbol{x}) = \frac{m_0(\boldsymbol{x})}{m_1(\boldsymbol{x})} = \frac{\prod_{i=1}^{k+1} ({\theta_{0i}}^{x_i}) \prod_{i=1}^{k+1}[\Gamma(\alpha_i)]\Gamma[\sum_{i=1}^{k+1}(\alpha_i + x_i)]}{\Gamma(\sum_{i=1}^{k+1} \alpha_i) \prod_{i=1}^{k+1} \Gamma(\alpha_i + x_i)}$
- Bayesian model selection: Binomial ∧ Beta model
  - Likelihood: $x \sim \text{Bin}(N, \theta)$
  - Hypotheses: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$
  - Prior: $\pi(\theta|H_0) = 1_{\theta_0}(\theta)$ and $\theta|H_1 \sim \text{Beta}(a, b)$
    - (i) $(a, b) = (1, 1)$: Uniform prior
    - (ii) $(a, b) = (c\,\theta_0, c - c\,\theta_0)$, $c > 0$: Beta prior centred on θ_0
  - Marginal density: $m_i(x) = \int_0^1 f(x|\theta)\pi(\theta|H_i)d\theta,\ i = 0, 1$
    - (i) $m_0(x) = \binom{N}{x}{\theta_0}^x(1 - \theta_0)^{N-x}$ (ii) $m_1(x) = \binom{N}{x}\frac{B(x+a, n-x+b)}{B(a,b)}$
  - Bayes factor: $B_{01}(x) = \frac{m_0(x)}{m_1(x)} = \frac{\theta_0^x(1-\theta_0)^{N-x}\,\Gamma(a)\,\Gamma(b)\,\Gamma(n+a+b)}{\Gamma(a+b)\,\Gamma(n+b-x)\,\Gamma(x+a)}$
- $\boldsymbol{\theta_0}$ represents agreement to either BL1 or BL2 digit probabilities. θ_0 represents agreement to one of BL1 or BL2 digit probabilities when considered individually.

Methodology Part II

- $\boldsymbol{x}$ represents the counts of first (or second) digits, and x the count of one particular digit.
- The choice of c defines how informative the priors are. Considering only $c \in \mathbb{N}$:
  - $c = 1$: the least informative priors centred on the null parameter values
  - $c = 22$ for BL1 and $c = 12$ for BL2: the least informative uni-modal priors centred on the null parameter values
- Prior model probabilities: $P(H_0) = \pi_0$ and $P(H_1) = 1 - \pi_0$
- Posterior model probabilities: $P(H_0|\text{data}) = \left(1 + \frac{1-\pi_0}{\pi_0} B_{01}(\text{data})^{-1}\right)^{-1}$
  - Berger and Sellke (1987) consider $\pi_0 = \frac{1}{2}$ to be the objective choice.
- P-value Calibration (Sellke, Bayarri, and Berger, 2001):
  - Let p_{obs} be a p-value on a classical test statistic. The lower bound on the Bayes factor in favour of H_0 is: $\underline{B}_{01}(p_{obs}) = -e\, p_{obs} \ln(p_{obs})$ if $p_{obs} < \frac{1}{e}$ and $\underline{B}_{01}(p_{obs}) = 1$ otherwise.
  - With $\pi_0 = \frac{1}{2}$, the lower bound on $P(H_0|\text{data})$ is: $\underline{P}(H_0|p_{obs}) = \left(1 + \underline{B}_{01}(p_{obs})^{-1}\right)^{-1}$
  - We will use this calibration on p-values from χ^2-tests and Nigrini's (2012) z-tests.

Data

The data was downloaded from the Eurostat website (Data/Database by themes/Economy and finance/Government statistics/Government finance statistics/Government deficit and debt/Government deficit/surplus, debt and associated data). Each countries' sample consists in the aggregation of all the numbers from the 38 tables in this category. This category of data is related to public deficit and public debt, which are variables that are important to the Stability and Growth Pact criteria. Rauch, Göttsche, Brähler, and Engel (2011) warn that macroeconomic statistics can be used by governments to portray a more favourable picture of their countries economic situation. The pressure for European Union's governments to comply with the Stability and Growth Pact criteria is an additional incentive for them to manipulate macroeconomic statistics. We included data from 1999 to 2015, and only countries that joined the Eurozone prior to 2006 were considered so that at least 10 years of data were available. Our samples meet the conditions that according to Hill (1995) typically result in the emergence of Benford's law.

Results: Multinomial ∧ Dirichlet Model & χ^2-Test

	BL1						BL2					
Country	N	$P(H_0\|\boldsymbol{x})$ $\boldsymbol{\alpha}=\mathbf{1}$	$\boldsymbol{\alpha}=\boldsymbol{\theta_0}$	$\boldsymbol{\alpha}=22\,\boldsymbol{\theta_0}$	$\underline{P}(H_0\|\boldsymbol{x})$	p_{obs}	N	$P(H_0\|\boldsymbol{x})$ $\boldsymbol{\alpha}=\mathbf{1}$	$\boldsymbol{\alpha}=\boldsymbol{\theta_0}$	$\boldsymbol{\alpha}=12\,\boldsymbol{\theta_0}$	$\underline{P}(H_0\|\boldsymbol{x})$	p_{obs}
Austria	619	0.001	0.988	0.000	0.000	0.000	614	1.000	1.000	1.000	0.357	0.082
Belgium	604	0.016	1.000	0.000	0.000	0.000	604	1.000	1.000	1.000	0.481	0.236
Finland	605	0.919	1.000	0.200	0.000	0.000	605	1.000	1.000	1.000	0.331	0.068
France	600	0.039	1.000	0.001	0.000	0.000	600	1.000	1.000	0.999	0.069	0.005
Germany	612	0.942	1.000	0.113	0.002	0.000	610	1.000	1.000	1.000	0.453	0.175
Greece	629	0.885	1.000	0.067	0.002	0.000	627	1.000	1.000	1.000	0.152	0.016
Ireland	616	0.004	0.997	0.000	0.000	0.000	616	1.000	1.000	1.000	0.500	0.387
Italy	625	0.945	1.000	0.110	0.002	0.000	608	1.000	1.000	1.000	0.346	0.075
Luxembourg	602	0.000	0.000	0.000	0.000	0.000	602	1.000	1.000	0.999	0.089	0.007
Netherlands	596	1.000	1.000	0.986	0.105	0.009	596	1.000	1.000	1.000	0.313	0.060
Portugal	617	0.000	0.000	0.000	0.000	0.000	617	1.000	1.000	1.000	0.461	0.189
Spain	535	0.808	1.000	0.018	0.002	0.000	530	1.000	1.000	1.000	0.500	0.855
Pooled Sample	7260	0.999	1.000	0.912	0.000	0.000	7229	1.000	1.000	1.000	0.317	0.061

For data, code, replicability, and supplementary materials see https://github.com/pedro-teles-fonseca/bayes-digit-analysis-warwick

Results: Binomial ∧ Beta Model & Z-Test

Austria BL1

Digit	$P(H_0\|x)$ $a=1$, $b=1$	$a=\theta_0$, $b=1-\theta_0$	$a=22\,\theta_0$, $b=22-22\,\theta_0$	$\underline{P}(H_0\|x)$	p_{obs}
1	0.542	0.618	0.232	0.170	0.019
2	0.951	0.966	0.813	0.500	0.493
3	0.954	0.962	0.782	0.500	0.406
4	0.194	0.290	0.059	0.018	0.001
5	0.016	0.030	0.005	0.000	0.000
6	0.973	0.978	0.844	0.500	0.740
7	0.001	0.000	0.000	0.000	0.000
8	0.947	0.941	0.647	0.462	0.191
9	0.968	0.976	0.812	0.500	0.422

Belgium BL1

Digit	$P(H_0\|x)$ $a=1$, $b=1$	$a=\theta_0$, $b=1-\theta_0$	$a=22\,\theta_0$, $b=22-22\,\theta_0$	$\underline{P}(H_0\|x)$	p_{obs}
1	0.003	0.003	0.001	0.000	0.000
2	0.028	0.047	0.009	0.000	0.000
3	0.008	0.014	0.003	0.000	0.000
4	0.963	0.967	0.799	0.500	0.489
5	0.940	0.941	0.674	0.471	0.210
6	0.960	0.961	0.749	0.499	0.334
7	0.977	0.980	0.850	0.500	0.934
8	0.957	0.969	0.777	0.496	0.301
9	0.920	0.903	0.515	0.401	0.113

Ireland BL1

Digit	$P(H_0\|x)$ $a=1$, $b=1$	$a=\theta_0$, $b=1-\theta_0$	$a=22\,\theta_0$, $b=22-22\,\theta_0$	$\underline{P}(H_0\|x)$	p_{obs}
1	0.894	0.930	0.696	0.460	0.186
2	0.957	0.967	0.821	0.500	0.599
3	0.772	0.782	0.355	0.275	0.045
4	0.015	0.028	0.005	0.000	0.000
5	0.453	0.401	0.090	0.111	0.010
6	0.055	0.097	0.015	0.000	0.000
7	0.948	0.945	0.672	0.472	0.213
8	0.872	0.843	0.393	0.330	0.067
9	0.802	0.743	0.257	0.256	0.039

Luxembourg BL1

Digit	$P(H_0\|x)$ $a=1$, $b=1$	$a=\theta_0$, $b=1-\theta_0$	$a=22\,\theta_0$, $b=22-22\,\theta_0$	$\underline{P}(H_0\|x)$	p_{obs}
1	0.000	0.000	0.000	0.000	0.000
2	0.046	0.048	0.010	0.018	0.001
3	0.921	0.930	0.662	0.460	0.187
4	0.966	0.974	0.834	0.500	0.663
5	0.000	0.000	0.000	0.000	0.000
6	0.966	0.974	0.818	0.500	0.494
7	0.955	0.954	0.708	0.489	0.264
8	0.894	0.873	0.450	0.364	0.086
9	0.977	0.978	0.830	0.500	0.690

Portugal BL1

Digit	$P(H_0\|x)$ $a=1$, $b=1$	$a=\theta_0$, $b=1-\theta_0$	$a=22\,\theta_0$, $b=22-22\,\theta_0$	$\underline{P}(H_0\|x)$	p_{obs}	Digit	$P(H_0\|x)$ $a=1$, $b=1$	$a=\theta_0$, $b=1-\theta_0$	$a=22\,\theta_0$, $b=22-22\,\theta_0$	$\underline{P}(H_0\|x)$	p_{obs}
1	0.000	0.000	0.000	0.000	0.000	6	0.974	0.979	0.848	0.500	0.848
2	0.001	0.001	0.000	0.000	0.000	7	0.881	0.862	0.435	0.349	0.077
3	0.964	0.971	0.828	0.500	0.662	8	0.972	0.978	0.834	0.500	0.591
4	0.001	0.001	0.000	0.000	0.000	9	0.979	0.982	0.853	0.500	0.959
5	0.972	0.976	0.834	0.500	0.726	-	-	-	-	-	-

Conclusion

The irreconcilability of classical and Bayesian measures of evidence in precise null hypothesis testing (Berger and Sellke, 1987) arises in digit analysis. Even lower bounds on $P(H_0|\text{data})$, which are biased towards H_1, often provide much more evidence in favour of H_0 than what p-values seem to suggest. One incurring in the p-value fallacy (Goodman, 1999) will typically overestimate the practical importance of an observed deviation from Benford's law. Classical tests of fixed dimension are of limited usefulness in digit analysis: they attain high levels of power in large samples and hence tiny deviations from Benford's law, without practical importance, are likely to be statistically significant. Conclusions drawn from classical tests, and agreement with Bayesian model selection, are sensible to an arbitrarily chosen evidence threshold. It can therefore be misleading to draw sharp conclusions based solely on statistical significance. Conclusions based solely on p-values, or obtained with the conventional 0.05 threshold, are likely to underestimate the evidence in favour of Benford's law.

References

Benford, Frank (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society* **78**(4), 551–572.

Berger, James O. and Thomas Sellke (1987). Testing a point null hypothesis: the irreconcilability of p values and evidence. *Journal of the American Statistical Association* **82**(397), 112–122.

Goodman, S. N. (1999). Toward evidence-based medical statistics 1: the p value fallacy. *Annals of Internal Medicine* **130**(12), 995–1004.

Hill, Theodore P. (1995). A statistical derivation of the significant-digit law. *Statistical science* **10**(4), 354–363.

Ley, Eduardo (1996). On the peculiar distribution of the US stock indexes' digits. *The American Statistician* **50**(4), 311–313.

Newcomb, Simon (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics* **4**(1), 39–40.

Nigrini, Mark (2012). *Benford's law: applications for forensic accounting, auditing, and fraud detection.* John Wiley & Sons.

Pericchi, Luis and David Torres (2011). Quick anomaly detection by the Newcomb—Benford law, with applications to electoral processes data from the USA, Puerto Rico and Venezuela. *Statistical Science* **26**(4), 502–516.

Rauch, Bernhard, Max Göttsche, Gernot Brähler, and Stefan Engel (2011). Fact and fiction in EU-governmental economic data. *German Economic Review* **12**(3), 243–255.

Sellke, Thomas, M. J. Bayarri, and James O. Berger (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician* **55**(1), 62–71.

Wasserstein, Ronald L. and Nicole A. Lazar (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician* **70**(2), 129–133.