Chains of distributions and Benford's Law

Dennis Jang (Dennis_Jang@brown.edu)
Jung Uk Kang (Jung_Uk_Kang@brown.edu)
Alex Kruckman (Alex_Kruckman@brown.edu)
Jun Kudo (Jun_Kudo@brown.edu)
Steven J. Miller (sjmiller@math.brown.edu)

Mathematics Department, Brown University

http://www.math.brown.edu/∼sjmiller/197

Workshop on Theory and Applications of Benford's Law
Santa Fe, NM, December 2007

Problem Outline

- Alex Kossovsky conjectured that many chains of distributions approach Benford's law.

Problem Outline

- Alex Kossovsky conjectured that many chains of distributions approach Benford's law.
- Consider $X_1 \sim \mathrm{Unif}(0, k)$, $X_2 \sim \mathrm{Unif}(0, X_1)$, ..., $X_n \sim \mathrm{Unif}(0, X_{n-1})$.

Problem Outline

- Alex Kossovsky conjectured that many chains of distributions approach Benford's law.
- Consider $X_1 \sim \mathrm{Unif}(0, k)$, $X_2 \sim \mathrm{Unif}(0, X_1)$, ..., $X_n \sim \mathrm{Unif}(0, X_{n-1})$.
- If $f_{n,k}(x_n)$ is the probability density for X_n, then

$$f_{n,k}(x_n) = \begin{cases} \frac{\log^{n-1}(k/x_n)}{k\Gamma(n)} & \text{if } x_n \in [0, k] \\ 0 & \text{otherwise.} \end{cases}$$

Problem Outline

- Alex Kossovsky conjectured that many chains of distributions approach Benford's law.
- Consider $X_1 \sim \mathrm{Unif}(0, k)$, $X_2 \sim \mathrm{Unif}(0, X_1)$, ..., $X_n \sim \mathrm{Unif}(0, X_{n-1})$.
- If $f_{n,k}(x_n)$ is the probability density for X_n, then

$$f_{n,k}(x_n) \ = \begin{cases} \frac{\log^{n-1}(k/x_n)}{k\Gamma(n)} & \text{if } x_n \in [0, k] \\ 0 & \text{otherwise.} \end{cases}$$

Theorem (JKKKM)

As $n \to \infty$ the distribution of digits of X_n rapidly tends to Benford's Law.

Uniform Density Example: $n = 10$ with 10,000 trials

Digit	Observed Probability	Expected Probability
1	0.298	0.301
2	0.180	0.176
3	0.127	0.125
4	0.097	0.097
5	0.080	0.079
6	0.071	0.067
7	0.056	0.058
8	0.048	0.051
9	0.044	0.046

Sketch of the proof

- First prove the claim for density $f_{n,k}$ by induction.
- Use Mellin Transforms and Poisson Summation to analyze probability.

Proof by Induction: Base Case: Calculating CDF

$$F_{2,k}(x_2) = \int_0^k \text{Prob}\left(X_{2,k} \in [0, x_2] | X_{1,k} = x_1\right) \text{Prob}(X_{1,k} = x_1) dx_1$$

Proof by Induction: Base Case: Calculating CDF

$$
\begin{aligned}
F_{2,k}(x_2) &= \int_0^k \mathrm{Prob}\left(X_{2,k} \in [0, x_2] | X_{1,k} = x_1\right) \mathrm{Prob}(X_{1,k} = x_1) dx_1 \\
&= \int_0^{x_2} \mathrm{Prob}(X_2 \in [0, x_2] | X_1 = x_1) \frac{dx_1}{k} \\
&\quad + \int_{x_2}^k \mathrm{Prob}(X_2 \in [0, x_2] | X_1 = x_1) \frac{dx_1}{k}
\end{aligned}
$$

Proof by Induction: Base Case: Calculating CDF

$$\begin{aligned} F_{2,k}(x_2) &= \int_0^k \mathrm{Prob}\left(X_{2,k} \in [0, x_2] | X_{1,k} = x_1\right) \mathrm{Prob}(X_{1,k} = x_1) dx_1 \\ &= \int_0^{x_2} \mathrm{Prob}(X_2 \in [0, x_2] | X_1 = x_1) \frac{dx_1}{k} \\ &\quad + \int_{x_2}^k \mathrm{Prob}(X_2 \in [0, x_2] | X_1 = x_1) \frac{dx_1}{k} \\ &= \int_0^{x_2} \frac{dx_1}{k} + \int_{x_2}^k \frac{x_2}{x_1} \frac{dx_1}{k} \;=\; \frac{x_2}{k} + \frac{x_2 \log(k/x_2)}{k}. \end{aligned}$$

Proof by Induction: Base Case: Calculating CDF

$$
\begin{aligned}
F_{2,k}(x_2) &= \int_0^k \mathrm{Prob}\left(X_{2,k} \in [0, x_2] | X_{1,k} = x_1\right) \mathrm{Prob}(X_{1,k} = x_1) dx_1 \\
&= \int_0^{x_2} \mathrm{Prob}(X_2 \in [0, x_2] | X_1 = x_1) \frac{dx_1}{k} \\
&\quad + \int_{x_2}^k \mathrm{Prob}(X_2 \in [0, x_2] | X_1 = x_1) \frac{dx_1}{k} \\
&= \int_0^{x_2} \frac{dx_1}{k} + \int_{x_2}^k \frac{x_2}{x_1} \frac{dx_1}{k} \;=\; \frac{x_2}{k} + \frac{x_2 \log(k/x_2)}{k}.
\end{aligned}
$$

Differentiating yields $f_{2,k}(x_2) = \frac{\log(k/x_2)}{k}$.

Further Comments

- Other distributions: exponential, one-sided normal.
- Weibull distribution: $f(x;\gamma) = \gamma x^{\gamma-1} \exp(-x^{\gamma})$.
- Further areas of research - Two parameter distribution, closed form for other single variable distributions