Lotka’s Law package

Alon Friedman, PhD
School of Information
A.J. Lotka

- The science of processing data for storage and retrieval, also known as informatics, investigates three components: the source of data, author productivity and word count.
- An important aspect for any academic or professional researcher is measuring the impact of their scientific productivity.
- In 1926, A. J. Lotka examined author publication productivity by looking at two conference proceedings in the fields of Chemistry and Physics.
- In his findings, he provided a predictable pattern for the relative contributions of a body of authors to a body of literature.
- He reported that 60% of authors make a single contribution during a given time period, 15% (1/2^2 times .60) of the authors publish two articles, and 7% (1/3^2 times .60) of the authors publish three articles.
- That means only 6% of the authors in a subject field and at a given time, produce more than ten articles.
Lotka's law of productivity

Proportional number of artists

Number of artworks created
Lotka’s Law

• Lotka’s Law is based on the formula:
  \( X^n Y = C \)

• \( X \) = the number of publications
  \( Y \) = the relative frequency of authors with \( X \) publications

\( n \) and \( C \) are constant depending on the specific field
(\( n \approx 2 \))
Finding the values of n and C

• Exponent of n

The formula of estimation of the exponent n

\[ n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2} \]

**N** = number of pairs of data

**X** = logarithm of x, i.e. number of publications

**Y** = logarithm of y, i.e. number of authors
Finding the values of n and C

• The constant C

If we accept Lotka’s conclusion that the proportion of all authors making a single contribution is about 60%, then the value of C can be computed by the simple formula $6/\pi^2$. However, if $n$ equals 2, C is the inverse of the summation of the infinite series: the limit of each equals to $\pi^2/6$.

• The formula:

$$c = \frac{1}{\sum_{1}^{p-1} \frac{1}{x^2} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2p^n} + \frac{n}{24(p-1)^{n-1}}}$$
Kolmogorov-Smirnov (K-S) test

- Pao (1985) suggests the K-S test, a goodness-of-fit statistical test, to assert that the observed author productivity distribution is not significantly different from a theoretical distribution:

\[
D = \max |F_0(x) - S_n(x)|
\]

\[F_0(x) = \text{theoretical cumulative frequency}\]

\[S_n(x) = \text{observed cumulative frequency}\]
The package

- The package holds 12 steps that allows us to calculate: C, N , K-S test and D valve (D = max/f0(x) - Sn (x))

**Step # 1**
CV <- function(Sums)

**Step # 2**
CVm <-
function (value, Sums)

**Step # 3**
LotkasC <- function(N)
{
  P <- 20
  increm <- c(1:(P-1))
  sum <- sum(1/increm^N)
  part1 <- sum
  part2 <- 1/((N-1)*(P*(N-1)))
  part3 <- 1/(2*(P^N))
  part4 <- N/(24*(P-1)^N)
  result <- (part1+part2+part3+part4)
  result <- 1/result
  return(result)
}

**Step 4**
LotkasN <- function(Sums,FullTable)
{
  N <- nrow(FullTable)
  lx <- Sums[3]
  ly <- Sums[4]
  xy <- Sums[5]
  x2 <- Sums[6]
  lx2 <- lx^2
  top <- (N*xy) - (lx*ly)
  bottom <- (N*x2) - (lx2)
  Nfinal <- top/bottom
  return(Nfinal)
}
The package

• **Step 5**
  
  ```
  results <- function(KTable)
  {
    percent <- function(x, digits = 2, format = "f", ...) {
      paste0(formatC(100 * x, format = format, digits = digits, ...), ",%")
    }
    value1 <- KTable[1:1,3:3]
    cat(KTable[1:1,2:2], " Authors made ", percent(value1))
    cat("\n")
    value1 <- KTable[2:2,3:3]
    cat(KTable[2:2,2:2], " Authors made ", percent(value1))
    cat("\n")
    value1 <- KTable[3:3,3:3]
    cat(KTable[3:3,2:2], " Authors made ", percent(value1))
  }
  ```

• **Step 6**

• **Step 7**

• **Step 8**

• **Step 9**

• **Step 10**

• **Step 11**

• **Step 12 Calculation of D value**

  ```
  LotkasXYfunction(Table){
    value <- (Table[,3] * Table[,4])
    return(value)
  }
  ```
What’s Next?

• The ability to handle multiple data sources.
• The ability to isolate single author vs. co-authors.
• New adaptations to Lotka's law.
Thank you

You can find the package at:
https://github.com/KCIV/LotkasLaw