Homework 1

STAT-GB.4310: Statistics for Social Data
Instructor: Patrick O. Perry

Due February 9, 2016

Theory

Suppose you have n texts, labeled $1, \ldots, n$. For $1 \leqslant i \leqslant n$, let w_i denote the length of text i, in words; let Y_i denote the number of times that a particular word appears in text i. Conditional on the text lengths, suppose that $Y_1, \ldots, Y_n$ are independent Poisson random variables with $E(Y_i) = w_i\lambda$ for some unknown rate parameter λ.

1. Give an expression for the likelihood function $L(\lambda)$, where

$$L(\lambda) = L(\lambda \mid y_1, \ldots, y_n) = \prod_{i=1}^{n} \Pr(Y_i = y_i \mid \lambda, w_i).$$

2. Define the log-likelihood function $l(\lambda) = \log L(\lambda)$. Give expressions for the first two derivatives, $l'(\lambda)$ and $l''(\lambda)$.

3. Compute an expression for $\hat{\lambda}$, the maximum likelihood estimator of λ. Use the fact that $\hat{\lambda}$ maximizes the log-likelihood $l(\cdot)$.

4. Show that $E(\hat{\lambda}) = \lambda$.

5. Compute $\mathrm{Var}(\hat{\lambda})$.

Application

Download `hw01.Rmd` and `federalist.json` from the course webpage. Fill in the missing code blocks to replicate Mosteller and Wallace's (1963) authorship analysis with the Poisson model. Answer the questions in `hw01.Rmd`. Print out and turn in the processed file, `hw01.html`.