

Exploration of E. coli protein copy numbers using Taylor's law

V2016-07-17 Rahul Dhodapkar <rahul.m.dhodapkar@gmail.com>

[Background](#)

[Taylor's Power Law](#)

[Analysis](#)

[Manipulation of Power Law](#)

[Plot and Fit](#)

[Open Questions](#)

[Potential biological implications of calculated Taylor's Law residuals?](#)

[Reproduce results with other protein copy number datasets.](#)

[Tie-ins with other research?](#)

[Cleanup Items](#)

[How does standard error affect the fidelity of these results?](#)

[Significant figures for final residuals output?](#)

Background

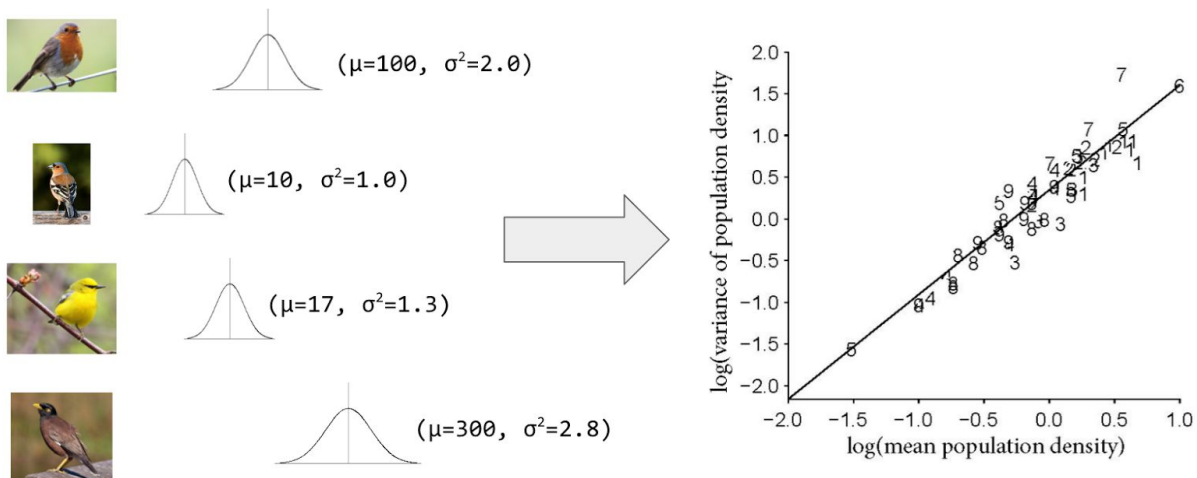
Working from the data provided in [Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells \(Taniguchi, Xie, et. al\)](#)¹, we would like to provide alternative statistical descriptions of the source data, to expose new properties of protein dynamics within the important model organism.

Taylor's Power Law

[Taylor's law](#)², also known as fluctuation scaling, is a model describing distributions of data often found in demography and ecology. The law describes a relationship between the means of a set of distributions, and the variances of those distributions commonly found among real populations. A diagram illustrating a high-level overview of the law is given below.

¹ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2922915/>

² https://en.wikipedia.org/wiki/Taylor%27s_law



NOTE: diagram serves to illustrate technique — exact figures not valid

Taylor's law in its most common form is given as the relation:

$$\sigma^2 = a\mu^b \quad [1]$$

Where σ is the standard deviation, and μ is the mean, while a and b are positive constants also known as the “residuals”. The values these residuals take on in a set of populations well-described by Taylor's law can be used to build hypotheses describing the source populations.

In our analysis here, we will be using Taylor's law to describe **populations of proteins** (protein copy number) in living cells, rather than populations of birds. Our goal is twofold, first to determine if Taylor's law is a good model for describing proteins in this way, and second to determine if any insights into protein dynamics can be made by examining the residuals of a good fit of protein number against Taylor's law.

Analysis

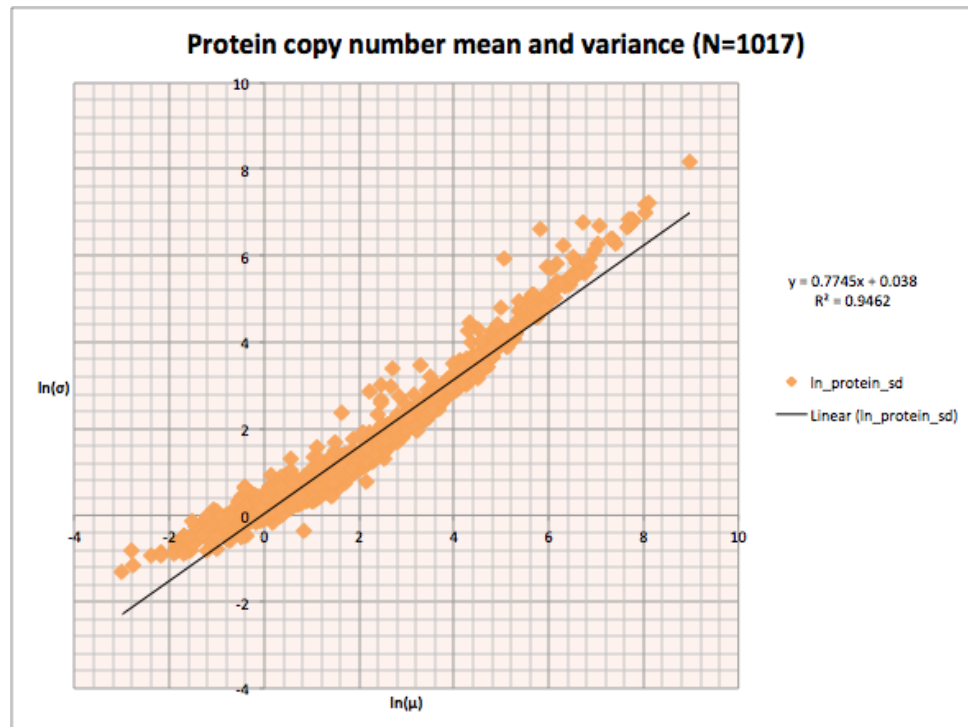
Manipulation of Power Law

To work with the power law, we perform some simple manipulations of the form:

$$\begin{aligned} \sigma^2 &= a\mu^b \\ \ln(\sigma^2) &= \ln(a\mu^b) \\ 2\ln(\sigma) &= \ln(a) + \ln(\mu^b) \\ \ln(\sigma) &= \ln(a)/2 + (b/2)\ln(\mu) \end{aligned} \quad [2]$$

Plot and Fit

Taking the data from supporting table S6 provided as [supplementary materials to the aforementioned paper](#)³, we perform a simple least-squares minimizing linear regression using Microsoft Excel.



Given [2] as defined above, we may now solve for the Taylor law residuals fairly easily as:

$$\begin{aligned} \ln(a)/2 &= 0.038 \\ a &= e^{0.076} = 1.079 & | \quad a &= 1.079 \quad | \\ b/2 &= 0.7745 \\ b &= 1.549 & | \quad b &= 1.549 \quad | \end{aligned}$$

We have now somewhat addressed one of our two original goals. We have determined with some cursory work that Taylor's law seems to be a reasonable model for describing protein copy number based on the raw data from Xie, Taniguchi. Moving forward, one would like to validate these results with other protein copy number datasets, and make some claims about what this relationship might mean about protein regulatory mechanisms in cells in general.

³ <http://science.sciencemag.org/content/suppl/2010/07/27/329.5991.533.DC1>

Open Questions

Potential biological implications of calculated Taylor's Law residuals?

Reproduce results with other protein copy number datasets?

Tie-ins with other research?

Develop mathematical model driven by Taylor's Law findings?

Cleanup Items

How does standard error affect the fidelity of these results?

Significant figures for final residuals output?