Deriving Book Sales From Rank

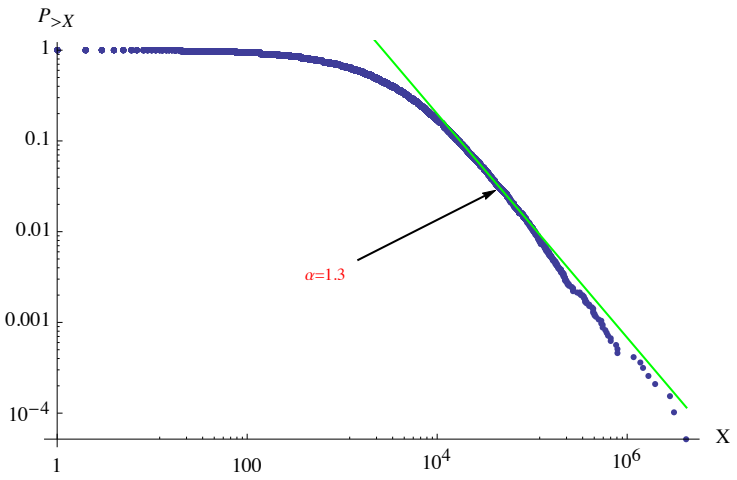
N. N. Taleb

I was priviledged to get access to a database with cumulative sales for editions in print that had at least one unit sold that particular week (that is, conditional of the specific edition being still in print). I fit a powerlaw with tail exponent $\alpha \simeq 1.3$ for the upper 10% of sales (graph), with N=30K. Using the Zipf variation for ranks of powerlaws, with r_x and r_y the ranks of book x and y, respectively, S_x and S_y the corresponding sales

$$\frac{S_x}{S_y} = \left(\frac{r_x}{r_y}\right)^{-\frac{1}{\alpha}}$$

So for example if the rank of x is 100 and y is 1000, x sells $\left(\frac{100}{1000}\right)^{-\frac{1}{1.3}} = 5.87$ times what y sells.

Note this is more robust in deriving the sales of the lower ranking edition $(r_y > r_x)$ because of inferential problems in the presence of fat-tails.



Log-Log Plot of the probability of exceeding X (book sales), and X

Technical Footnote: this works best for the top 10,000 books, but not quite the top 20 (because the tail is vastly more unstable). Further, the effective α for large deviations is lower than 1.3. But this method is robust as applied to rank.