6.3 MLE and Power Laws

If we have a power-law probability density, of index say β, then the normalized density is

$$\begin{aligned} \mathrm{prob}(s|\beta,\lambda) &= \tfrac{(\beta-1)s^{-\beta}}{\lambda^{1-\beta}} \quad && s \geq \lambda \\ &= 0 && \text{otherwise} \end{aligned}$$

assuming a lower limit λ, and also that the power law is steep enough for us to ignore any upper limit.
If we have N observations S_i then the log-likelihood is

$$\begin{aligned} \mathcal{L} &= N\log(\beta-1) - N(1-\beta)\log\lambda - \beta\textstyle\sum_i \log S_i \quad && S_1 \geq \lambda, S_2 \geq \lambda \ldots \\ &= \log 0 && \text{otherwise.} \end{aligned}$$

Differentiating with respect to β will give the result in the text. However, the likelihood appears to have no maximum in λ - it just rises as $N(\beta-1)\log\lambda$. Viewed as a function of λ, we see that the conditions $S_1 \geq \lambda, S_2 \geq \lambda \ldots$ mean that λ must be smaller than (or equal to) the smallest datum, if the likelihood is to be non-zero. It follows that the maximum likelihood is at

$$\hat{\lambda} = S_{\mathrm{min}}.$$

If we take a Bayesian perspective, then we may be interested in the maximum of the posterior density, sometimes abbreviated MAP. In this case, the prior on λ may matter. One interesting point arises. Suppose we had a flat prior, only non-zero above a lower cutoff at λ_0. If we had initially guessed a λ_0 that was actually bigger than the smallest datum we got in the experiment, the posterior probability for λ would be undefined! (It is 0/0.)
However there is an uncomfortable feeling here about a prior which may be flatly incompatible with the data. This problem has an affinity with the famous "taxicab problem" – see Jaynes 2003 ¶6.20
An important moral is: don't use priors which go to zero, unless there is a very good reason. If you do, the data may never be able to change your mind.