6.3 MLE and Power Laws

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

$$\text{prob}(s|\beta, \lambda) = \begin{cases} \frac{(\beta-1) s^{-\beta}}{\lambda^{1-\beta}} & s \geq \lambda \\ 0 & \text{otherwise} \end{cases}$$

assuming a lower limit $\lambda$, 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

$$L = N \log(\beta - 1) - N(1 - \beta) \log \lambda - \beta \sum \log S_i \quad S_1 \geq \lambda, S_2 \geq \lambda \ldots$$

= $\log 0$ otherwise.

Differentiating with respect to $\beta$ will give the result in the text. However, the likelihood appears to have no maximum in $\lambda$ - it just rises as $N(\beta - 1) \log \lambda$. Viewed as a function of $\lambda$, we see that the conditions $S_1 \geq \lambda, S_2 \geq \lambda \ldots$ mean that $\lambda$ 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

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

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 $\lambda$ may matter. One interesting point arises. Suppose we had a flat prior, only non-zero above a lower cutoff at $\lambda_0$. If we had initially guessed a $\lambda_0$ that was actually bigger than the smallest datum we got in the experiment, the posterior probability for $\lambda$ 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.