Joint distribution of estimators for log-likelihood and entropy

Anthony Ebert

29/08/2019

I was interested in why inclusion of entropy seems to improve the regularity of the KL divergence near the true parameter for the Erdos-Renyi model, so I looked at some much simpler examples.

Let's forget about networks temporarily and just think about discrete distributions such as the Poisson distribution and the negative binomial distribution. Set the mechanistic model parameters θ_q to equal the statistical model parameters θ.

$$
\begin{aligned}
KL[q(\cdot|\theta_q)||f(\cdot|\theta)] &= \sum_{y\in\mathcal{Y}} q(y|\theta_q)\log\frac{q(y|\theta_q)}{f(y|\theta)} \\
&= \sum_{y\in\mathcal{Y}} q(y|\theta_q)\log q(y|\theta_q) - \sum_{y\in\mathcal{Y}} q(y|\theta_q)\log f(y|\theta).
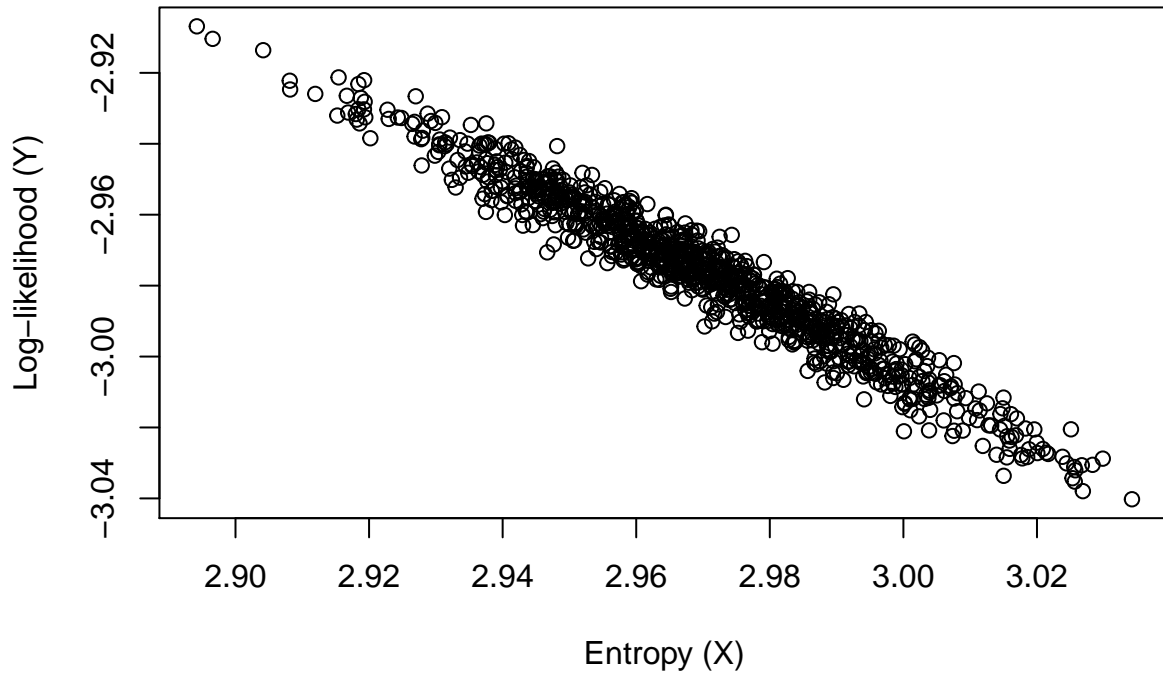\end{aligned}
$$

Let X be the first term (entropy) and Y be the second term (log-likelihood). What is the joint distribution (X, Y)?

X is calculated using the non-parameteric entropy estimator I have introduced previously and Y is calculated using the expression for f. The plots of X and Y used 1000 simulations for each point.

Poisson distribution
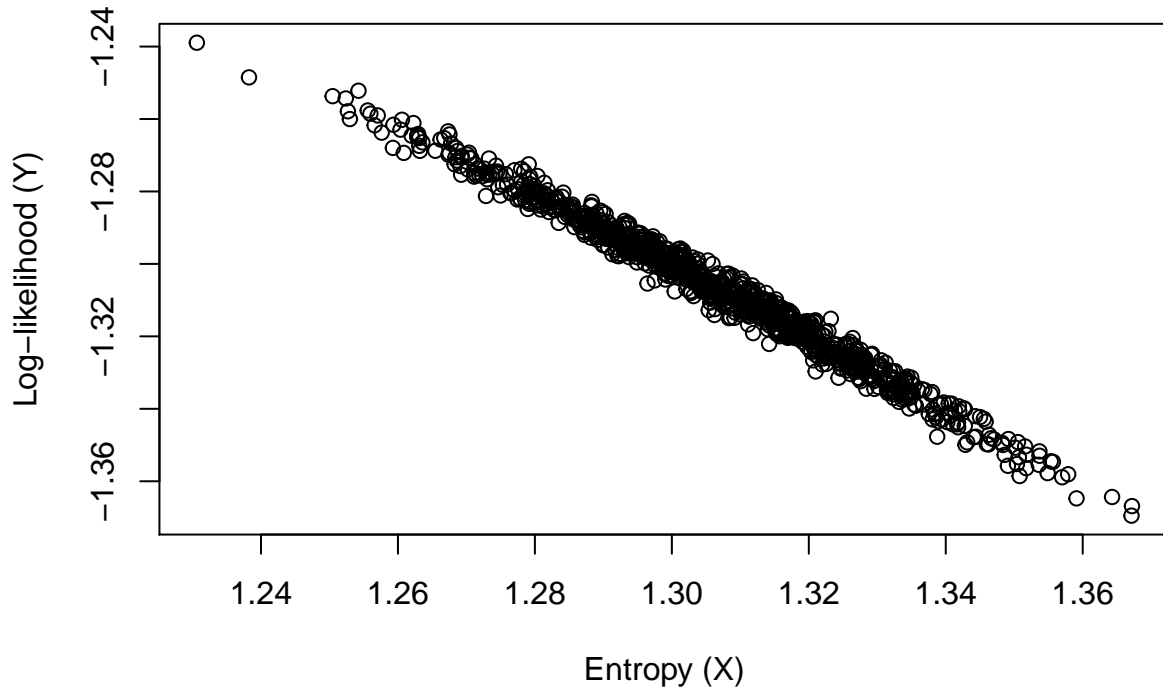
Negative binomial distribution

It seems like the slope is almost exactly -1. *With hindsight* this makes sense because the entropy just like a (negative) estimator for the log-likelihood. I'm sorry if this is obvious to you, it was a complete surpise for me.

I think some applications could include (if this hasn't already been known for 50 years):

- Zero variance control variates,
- Estimation of normalising constants - particularly for ERGM models, and obviously,
- Mapping mechanistic network models to statistical network models

Although these examples are uninteresting, I suspect that the estimator for entropy performs well even for high-dimensional, yet countable, sample spaces.