Report

Anthony Ebert

22/08/2019

We are interested in finding the closest mechanistic model $q(\cdot|\lambda)$, parameterised by λ to a known statistical model $p(\cdot|\theta)$, where θ is known.

$$\begin{aligned} KL[q(\cdot)||p(\cdot|\theta)] &= \sum_{y\in\mathcal{Y}} q(y)\log\frac{q(y)}{p(y|\theta)} \\ &= \log[z(\theta)] - \sum_{y\in\mathcal{Y}} q(y)\left[\sum_i \theta_i s_i(y)\right] + \sum_{y\in\mathcal{Y}} q(y)\log q(y) \end{aligned}$$

Entropy estimation

I use the following non-parameteric estimator of entropy (Vu, Yu, and Kass 2007):

$$\begin{aligned} \tilde{H} &:= -\sum_k \frac{\tilde{p}_k \log \tilde{p}_k}{1-(1-\tilde{p}_k)^n} \\ \tilde{p}_k &:= \hat{C}\hat{p}_k \\ \hat{C} &:= 1 - \frac{\#\{k|n_k = 1\}}{\sum_k n_k} \\ \hat{p}_k &:= \frac{n_k}{n} \end{aligned}$$

For instance

```
## [1] 0 1 1 2 0 2
```

```
## [1] 1.292132
```

```
## [1] 1.295627
```

```
## [1] 2 3 2 4 1 0
```

```
## [1] 1.63193
```

```
## [1] 1.625031
```

The relative importance of the likelihood and entropy

Attempt with preferential attachment model

The preferential attachment model has the following likelihood function:

$$P(k|\rho) = \frac{(\rho-1)\Gamma(k)\Gamma(\rho)}{\Gamma(k+\rho)}$$

In this case there is no simple set of summary statistics.

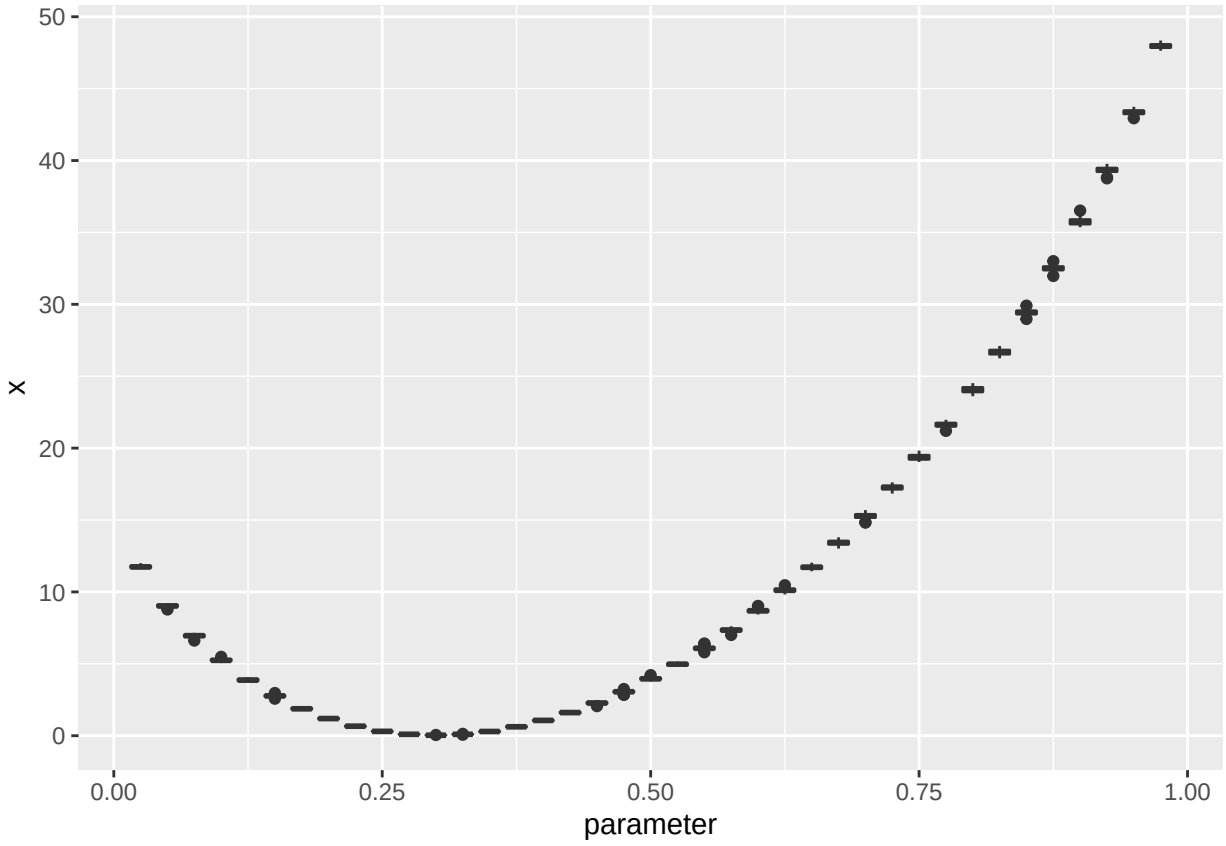


Figure 1: KL divergence calculation with entropy

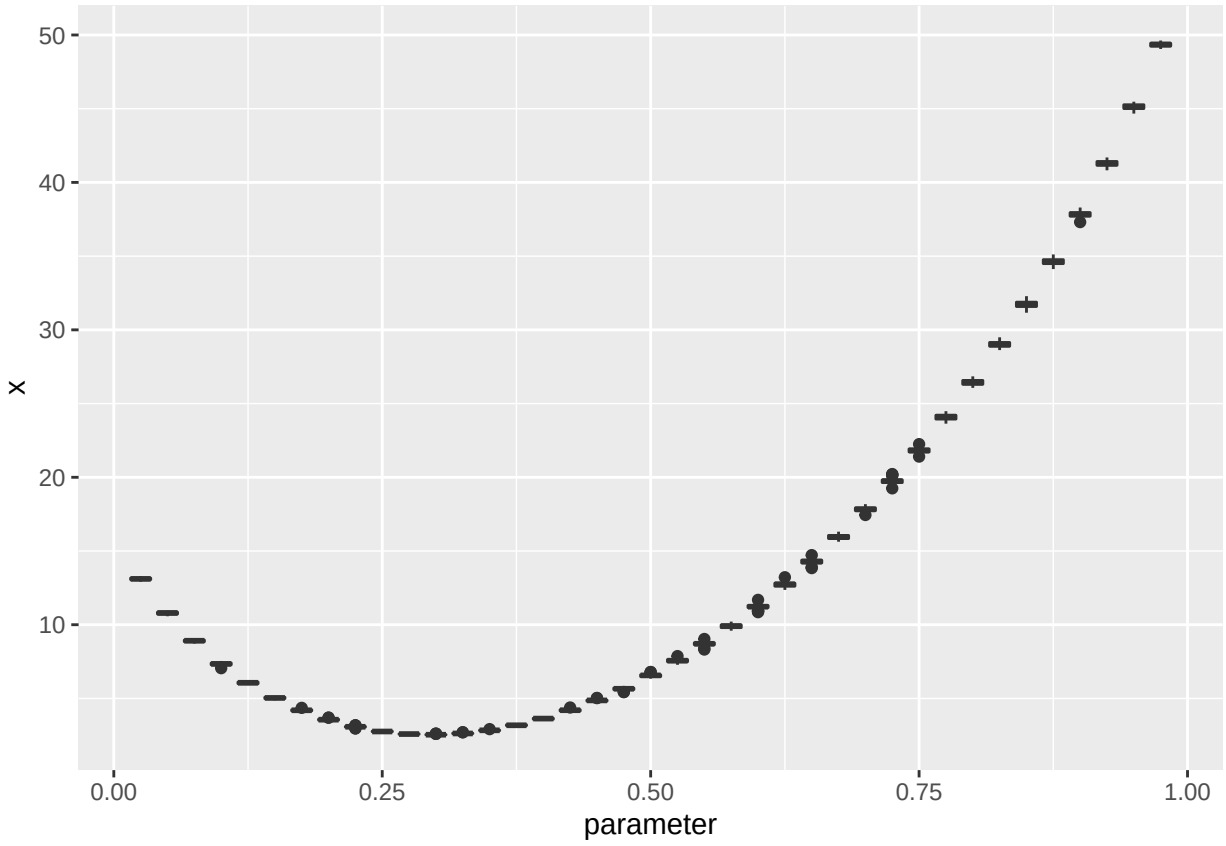


Figure 2: KL divergence calculation without entropy

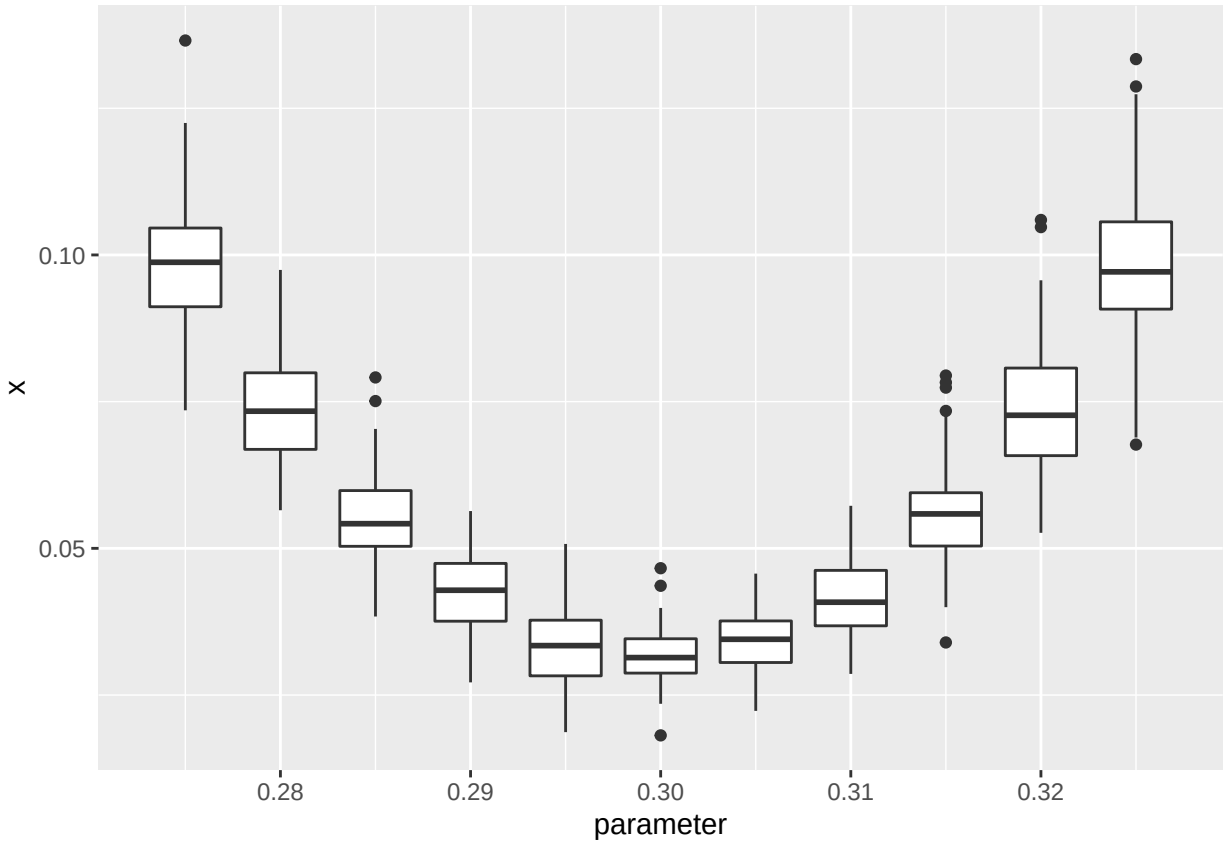


Figure 3: KL divergence calculation with entropy

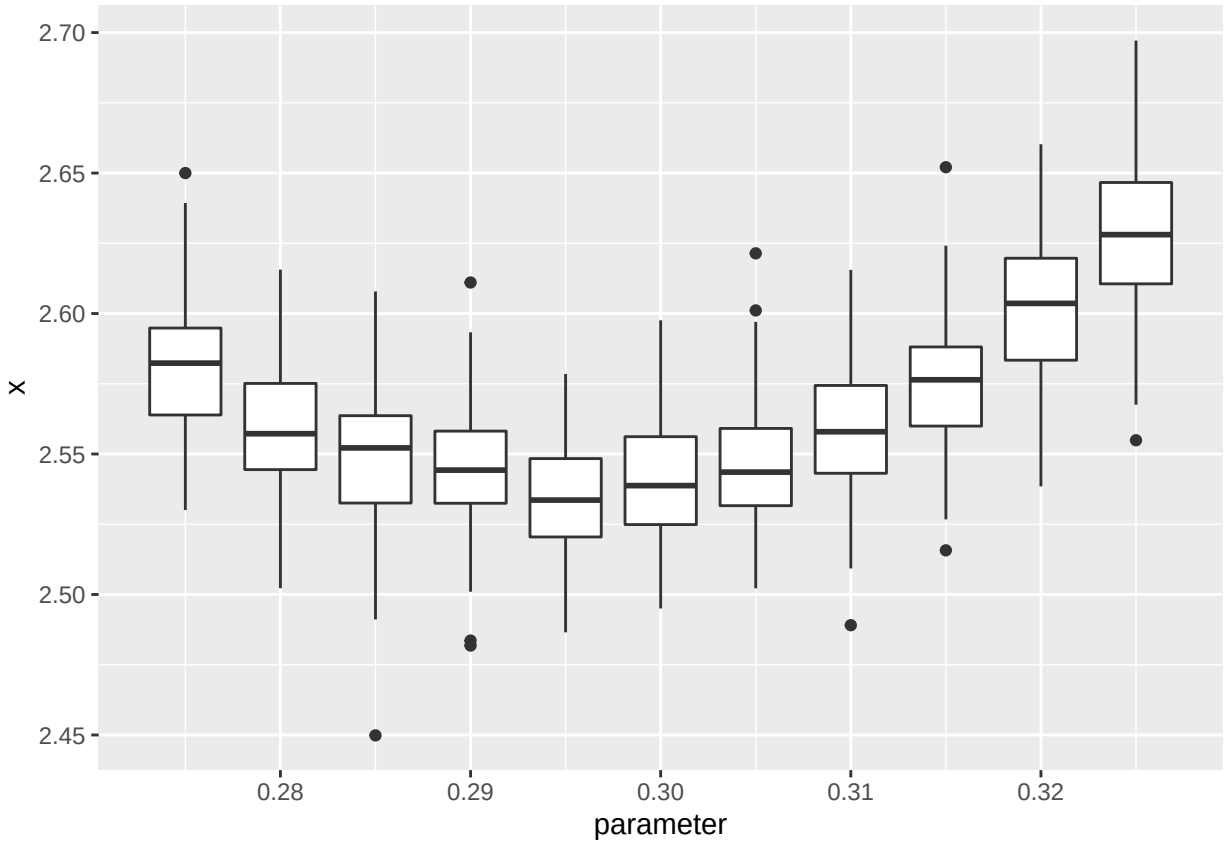
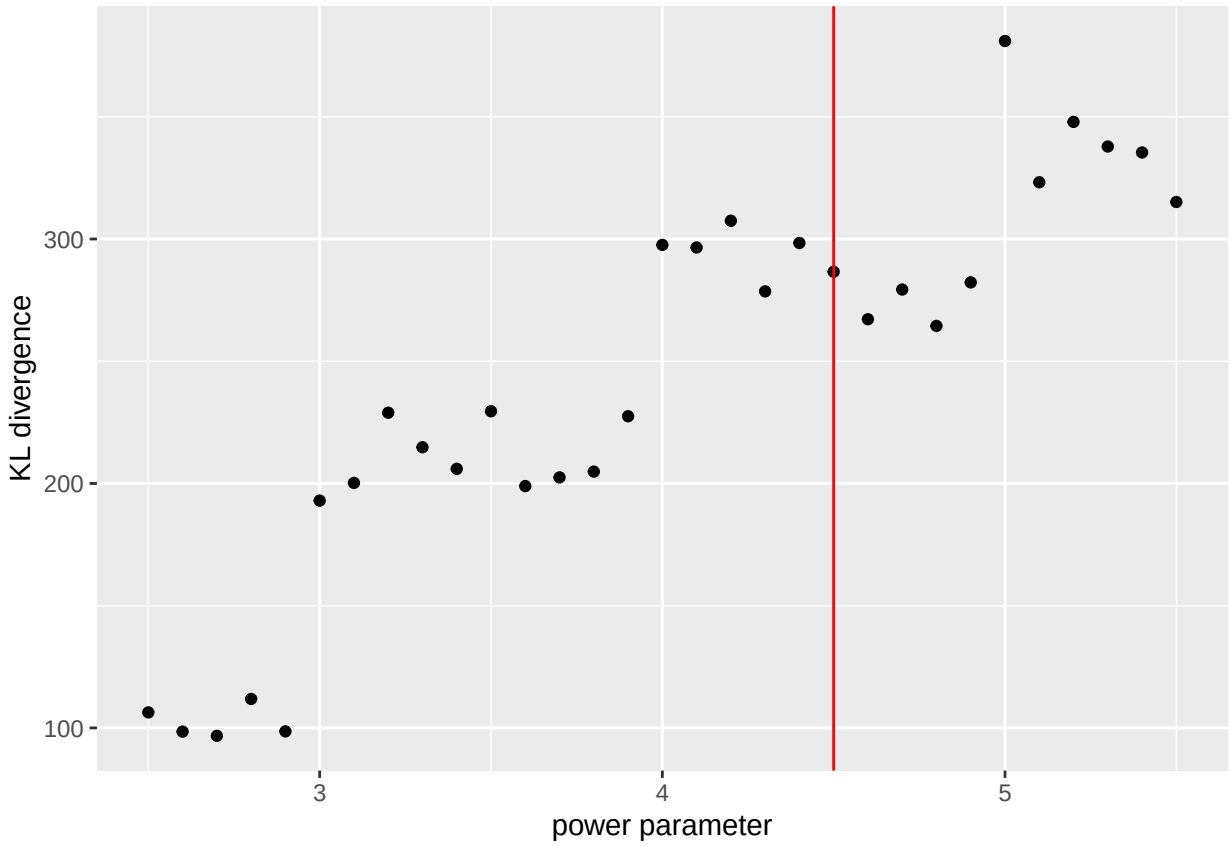


Figure 4: KL divergence calculation without entropy



Vu, Vincent Q, Bin Yu, and Robert E Kass. 2007. "Coverage-Adjusted Entropy Estimation." *Statistics in Medicine* 26 (21). Wiley Online Library: 4039–60.