NEURAL DATA ANALYSIS

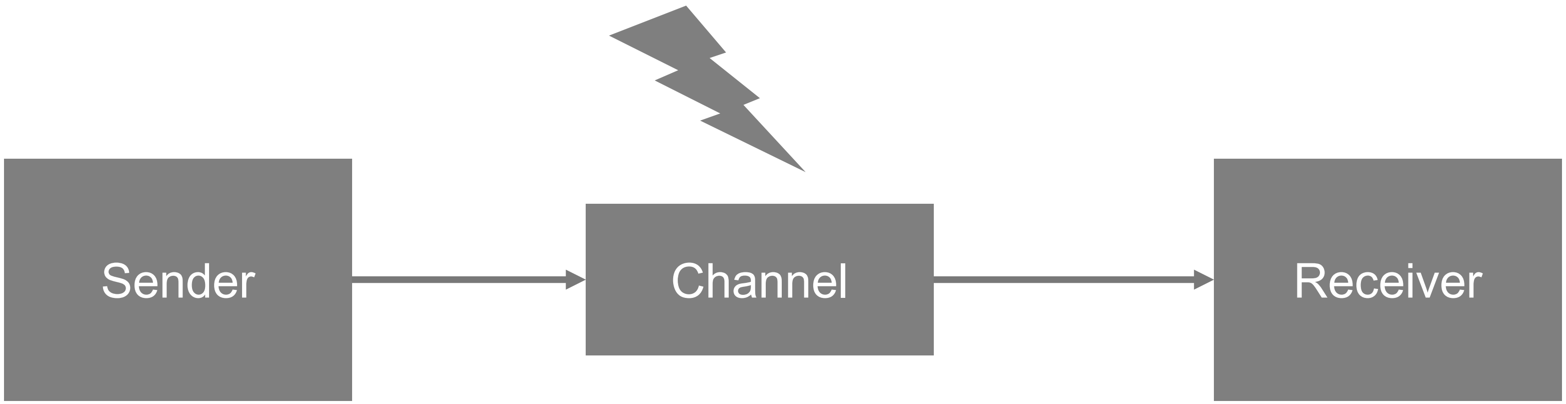
ALEXANDER ECKER, PHILIPP BERENS, MATTHIAS BETHGE

COMPUTATIONAL VISION AND NEUROSCIENCE GROUP

INFORMATION THEORY AND ENTROPY ESTIMATION

TASK 9

INFORMATION THEORY



X $\qquad$ Y

$p(X)$ $\qquad$ $p(Y)$

Stimulus $\qquad$ Spike train

UNCERTAINTY

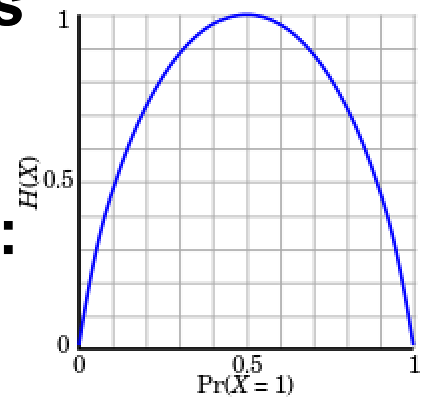
Entropy

Discrete!!!

$$H[X] = -\sum_{x \in X} p(x)\, log_2\, p(x)$$

Examples:

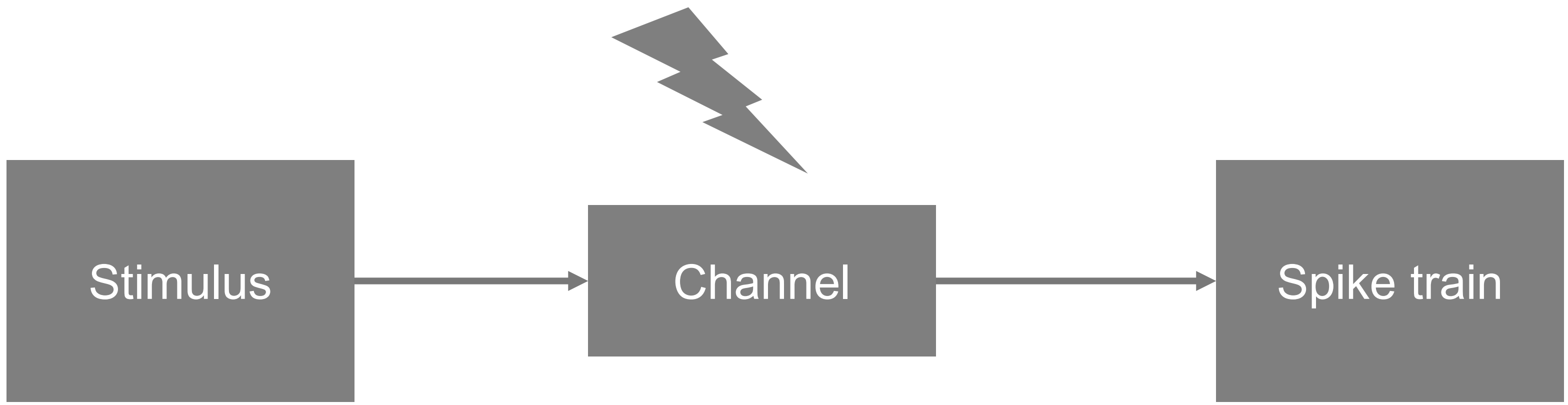
- **Entropy of a coin (bernoulli distribution)**
- **Entropy of K equally like outcomes**

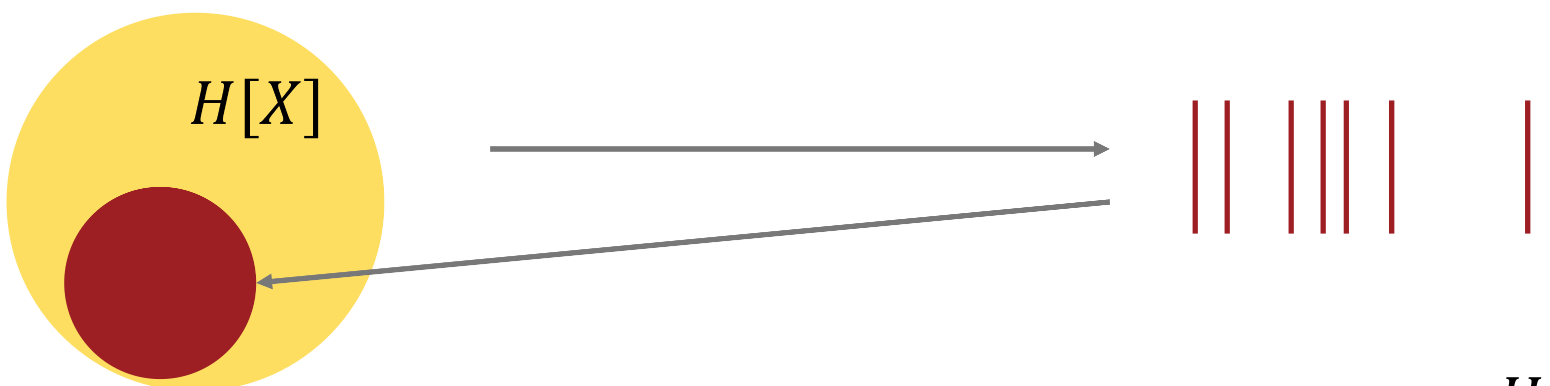


Average number of binary questions:

$$H[X] \leq n \leq H[X] + 1$$

CONDITIONAL ENTROPY





$$H[X|Y] = -\sum_{x} P(x,y) \log P(y|x)$$

$H[Y]$

$H[Y|X]$

MUTUAL INFORMATION

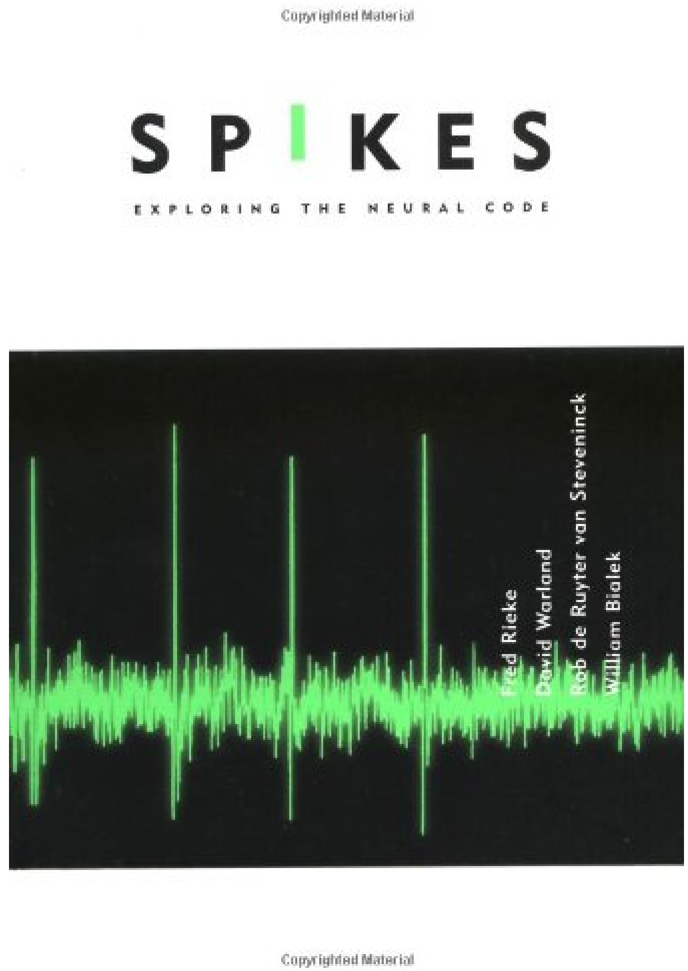
$$I[X,Y] = H[X] - H[X|Y]$$

Reduction in uncertainty about the stimulus after observing a spike train

$$I[X,Y] = H[Y] - H[Y|X]$$

Reduction in uncertainty about which spike train will be observed knowing the stimulus

INFORMATION THEORY IN NEUROSCIENCE



Quantitative framework

Ideal observer

Information need not be used by organism

Difficulty in estimation

ESTIMATING ENTROPY

Spike trains as short discrete sequences of spike or no spike

0101001100

$$x \in Z_2^{10}$$

M = 1024 states

$$p(X) = \frac{1}{M} = \frac{1}{1024}$$

$$H[X] = 10\ bits$$

ESTIMATING ENTROPY

Observe $x_1, x_2, x_3, \dots, x_N$

Count how often

$$f_1 = \#[x_i == 0000000000]$$
$$f_2 = \#[x_i == 0000000001]$$

$$\dots.$$

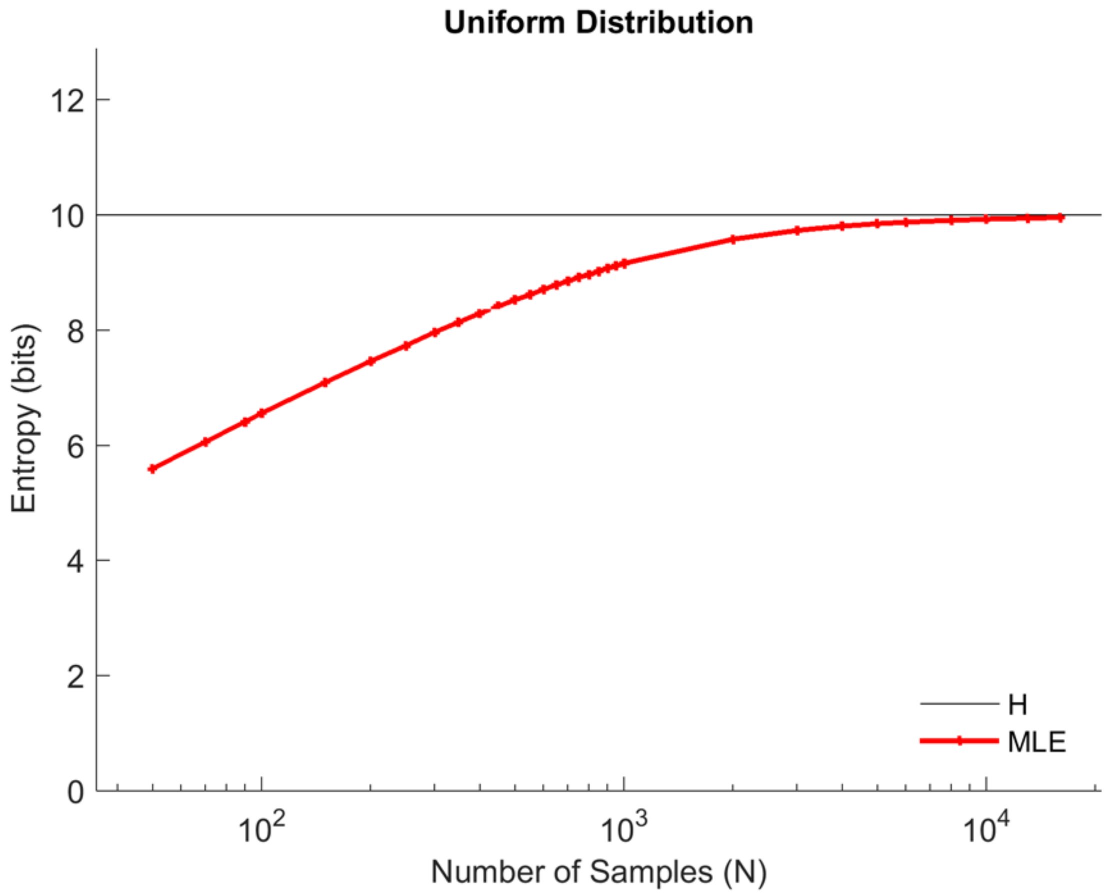
$$\hat{p}(X) = [\frac{f_1}{N}, \frac{f_2}{N}, \dots, \frac{f_{1024}}{N}]$$

Unobserved symbols!

Maximum likelihood estimator

$$\hat{H}_{ML} = -\sum_x \hat{p}(x) \, log \, \hat{p}(x)$$

BIAS OF PLUG-IN ESTIMATOR



EFFECT ON MUTUAL INFORMATION

$$I[X, Y] = H[Y] - H[Y|X]$$

- $H[Y]$: **for the overall spike train distribution** $P(Y)$ **we pool over all stimuli -> bias small**
- $H[Y|X]$: **for estimating the stimulus-conditional spike train distribution** $P(Y|X)$, **we have much less data -> bias large**
- **Overall: Overestimate** $I[X, Y]$**!**

MILLER MADDOW CORRECTION

One can show (Taylor expansion):

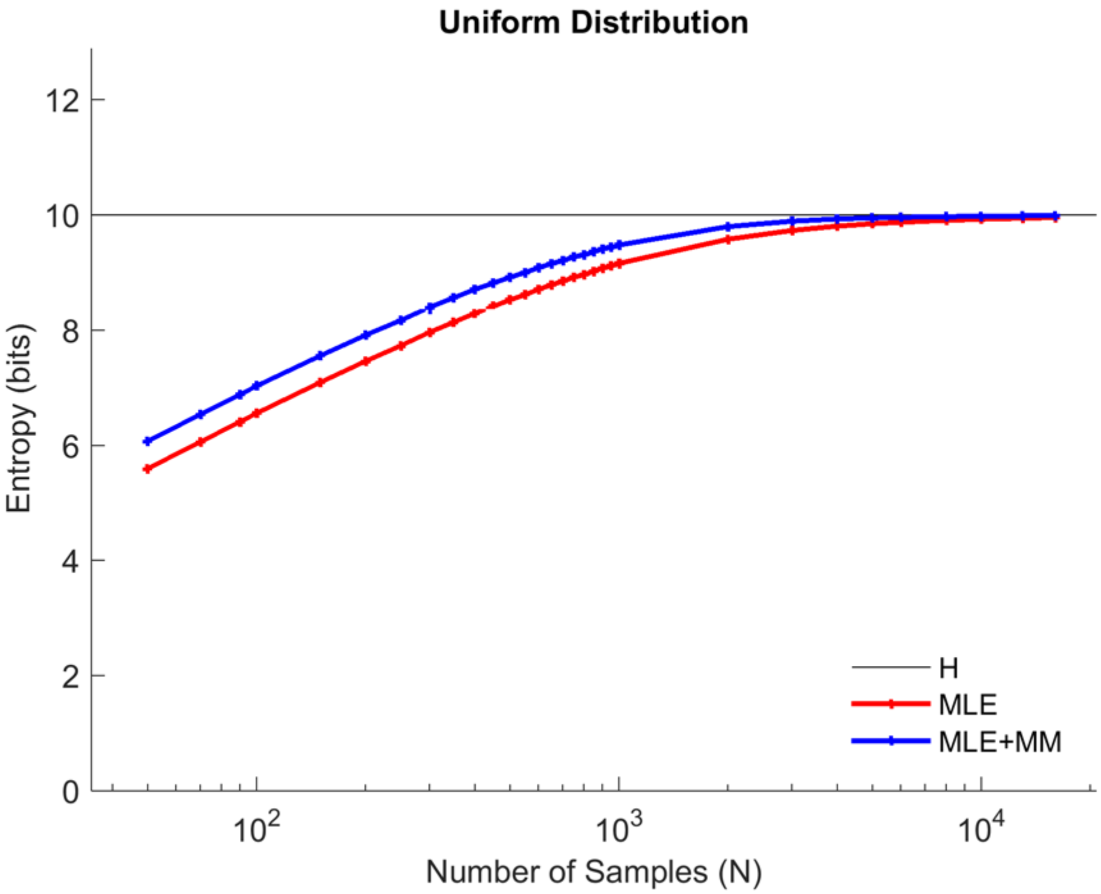
$$\boldsymbol{Bias} = -\frac{\boldsymbol{d}-\mathbf{1}}{\mathbf{2}\boldsymbol{n}} + \boldsymbol{O}(\boldsymbol{n}^{\mathbf{2}})$$

Bias-corrected estimator:

$$\boldsymbol{H}_{\boldsymbol{MM}} = \boldsymbol{H}_{\boldsymbol{ML}} + \frac{\widehat{\boldsymbol{d}}-\mathbf{1}}{\mathbf{2}\boldsymbol{n}}$$

$$\widehat{\boldsymbol{d}} = \#[\widehat{\boldsymbol{p}} > \mathbf{0}]$$

MM ESTIMATOR



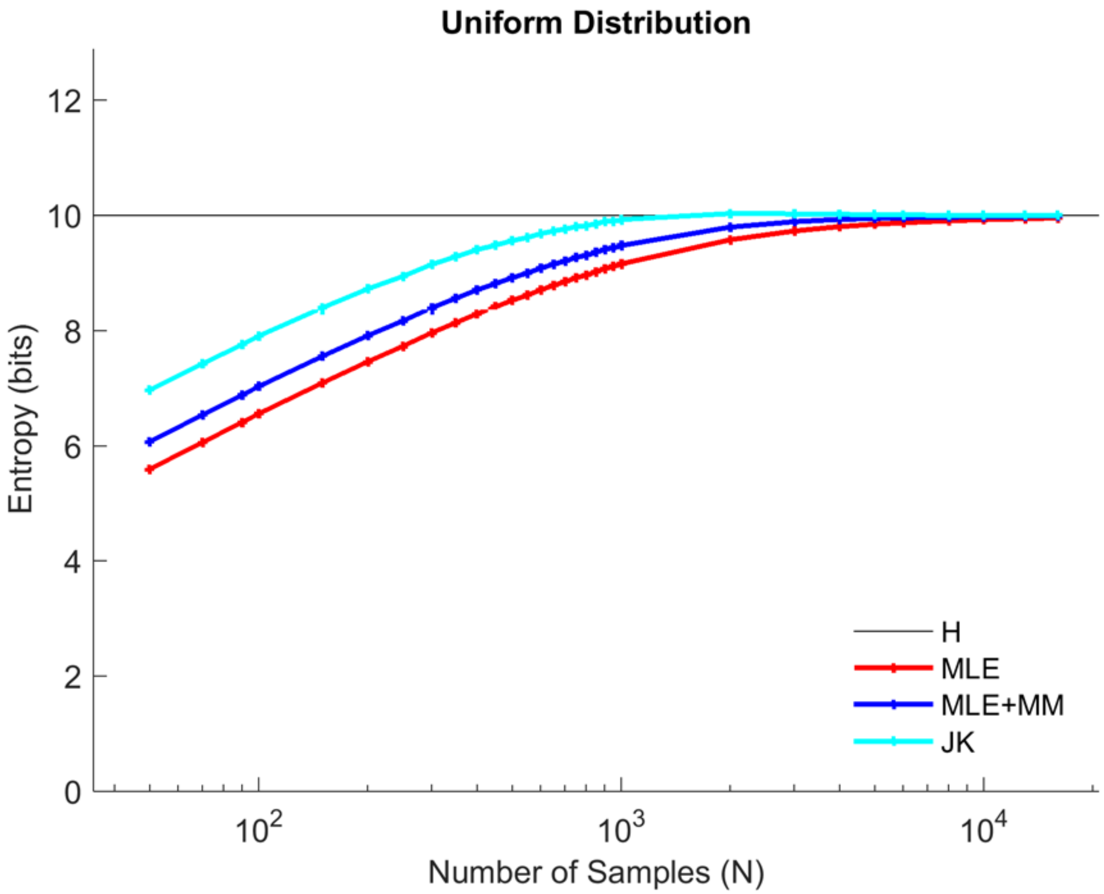
JACKKNIFE ESTIMATOR

Resampling method useful for variance or bias estimation

$$\hat{H}_{JK} = N\hat{H}_{ML} - (N-1)\,\hat{H}_{ML}^{(.)}$$
$$\hat{H}_{ML}^{(.)} = \langle H_{ML}^{\setminus \mathrm{i}} \rangle$$

Reduces bias by an order of magnitude

JACKNIFE ESTIMATOR



COVERAGE ADJUSTED ESTIMATOR

S is set of observed symbols

$$H[X] = -\sum_{x \in S} P(x) \log P(x) - \sum_{x \notin S} P(x) \log P(x)$$

Inflated summands
$x \in S: \widehat{P}(x) > P(x)$

Unobserved summands
$x \notin S: \widehat{P}(x) = 0$

Adjust for both

COVERAGE ADJUSTED ESTIMATOR

Shrink estimate of P for observed symbols:

$$C = 1 - \frac{\# f_i = 1}{N}$$

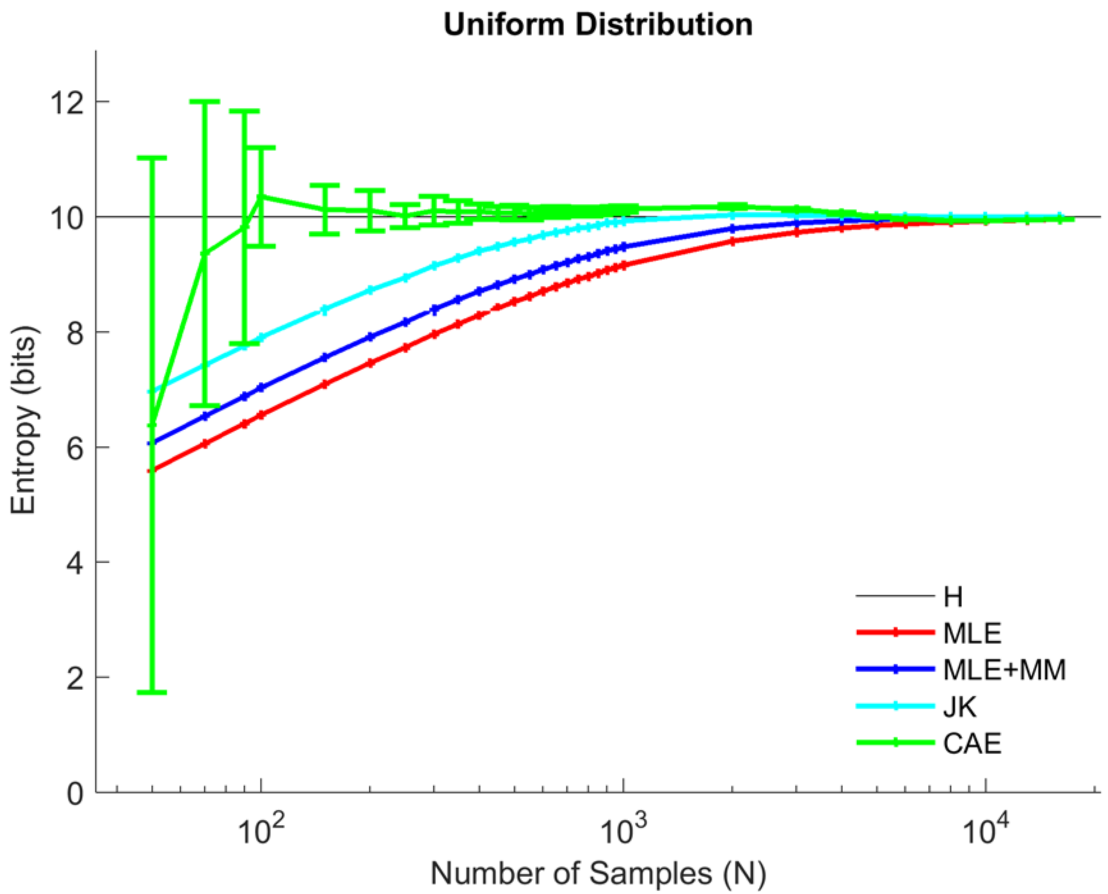
$$\hat{P}_C = \hat{P} * C$$

Inflate contribution for rare words:

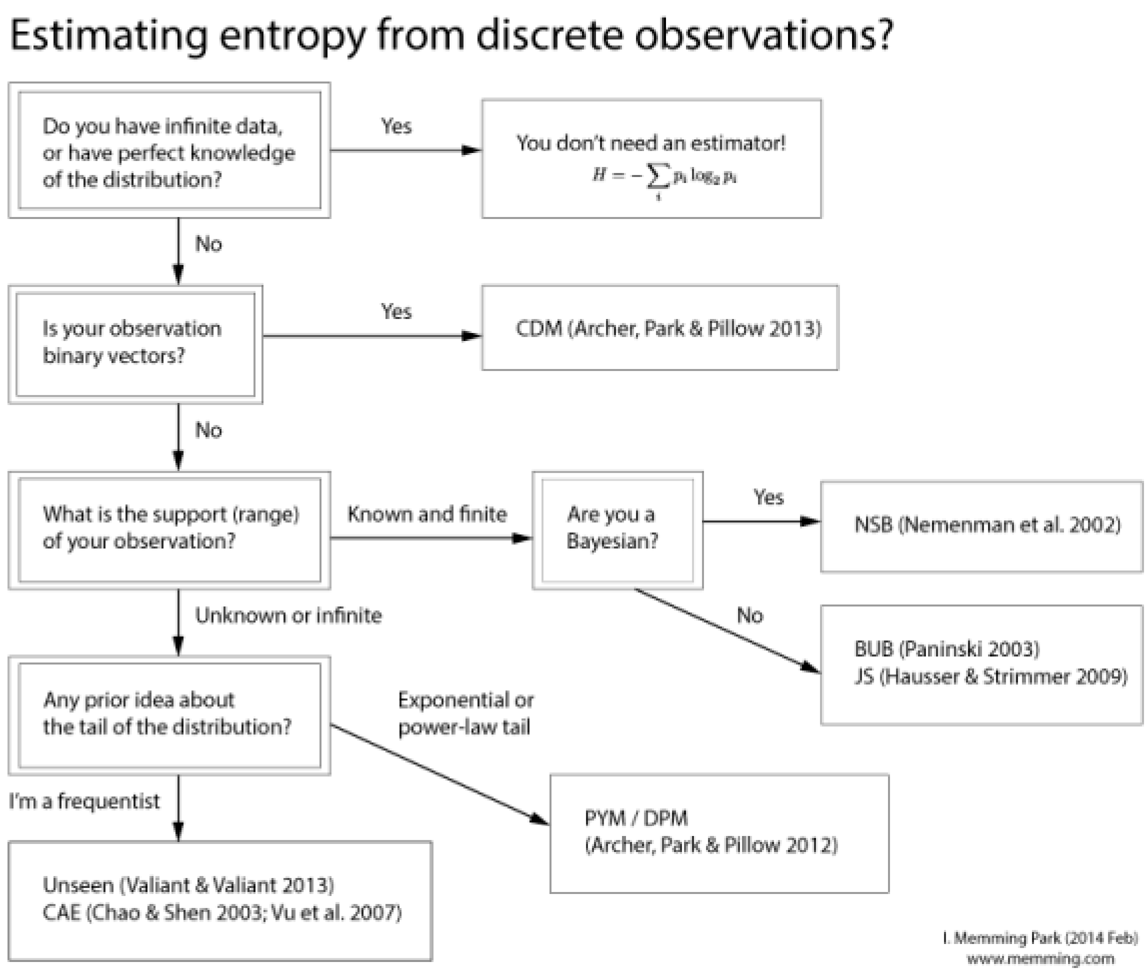
$$\hat{H}_{CA} = -\sum_x \frac{\hat{P}_C(x) \log \hat{P}_C(x)}{1 - \left(1 - \hat{P}_C(x)\right)^N}$$

Equivalent to regularized estimate of P

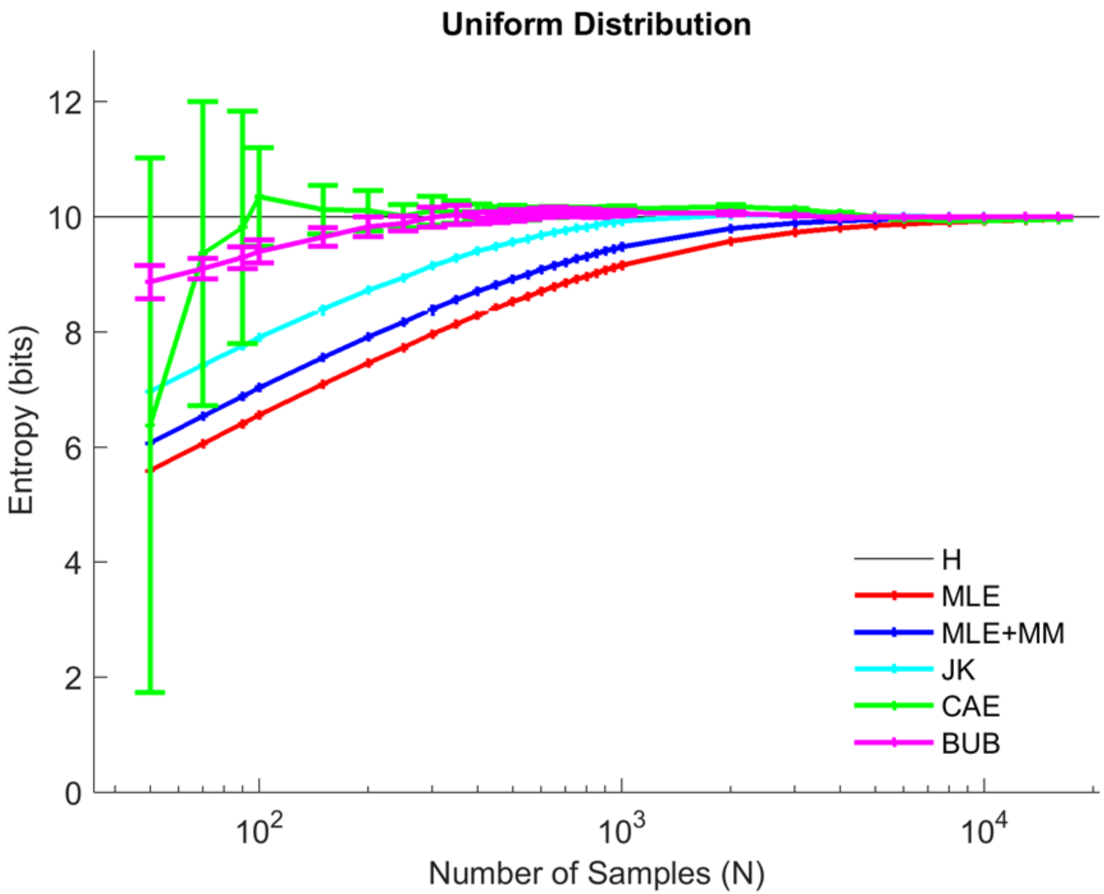
CAE ESTIMATOR



OTHER OPTIONS



BEST UPPER BOUNDS



LINKS

http://www.nowozin.net/sebastian/blog/estimating-discrete-entropy-part-1.html

http://www.nowozin.net/sebastian/blog/estimating-discrete-entropy-part-2.html

http://www.nowozin.net/sebastian/blog/estimating-discrete-entropy-part-3.html

https://memming.wordpress.com/2014/02/09/a-guide-to-discrete-entropy-estimators/

http://theory.stanford.edu/~valiant/papers/nips_full.pdf