Entropy estimation: coincidences, additivity, and uninformative priors

Ilya Nemenman CCS-3/CNLS, LANL; CDDMA, NMCIAS

nsb-entropy.sf.net



CNS'07, Toronto





Characterize coding without explicit decoding

$$S[x] = -\sum_{x} p(x) \log p(x), \qquad x = s, \{t_i\}$$
$$I[s, \{t_i\}] = \sum_{s \in t_i} p(s, \{t_i\}) \log \frac{p(s, \{t_i\})}{p(s)p(\{t_i\})}$$
$$= S[r] - S[r \mid s]$$

- Captures all dependencies (zero *iff* joint probabilities factorize)
- Reparameterization invariant
- Unique metric-independent measure of "how related"
- Because this it is the currency of sensory systems



Slide 2



Experiment design



Slide 3



Experiment design: probing precise spike timing





Problems

- Total of about 10-15 min of recordings (limited by stationarity of the outside world)
- At most 200 repetitions
- Stimulus correlation of 60ms: only 10000 independent samples (repeated or nonrepeated)
- Need to sample words of length 30 ms (behavioral) to 60 ms (stimulus) at resolution down to 0.2 ms (binary words of length up to ~100).

Undersampling!



Slide 5



Undersampling and entropy/MI estimation

Maximum likelihood estimation:





Undersampling and entropy/MI estimation



Fluctuations underestimate entropies and overestimate mutual informations.

(Need "smoothing")



Slide 7



Correct smoothing possible



 $S \leq \log N$

Incorrect smoothing -over- or underestimation.

Even refractory Poisson process at these T, τ has over 15-20 bits of entropy!

For estimation of entropy at *K*/*N*<<1 see: Grassberger 1989, 2003, Antos and Kontoyiannins 2002, Wyner and Foster 2003, Batu et al. 2002, Paninski 2003, Panzeri and Treves 1996, Strong et al. 1998. Not sure how to quantify Victor 97-02. Shlens et al is for *K*^*a*/*N*<<1, *a*<1.



Slide 8



What if S>logN?

- But there is hope (Ma, 1981) (*capture-recapture*):
- For uniform K-bin distribution the first coincidence occurs at

$$N_c \sim \sqrt{K} = \sqrt{2^S}$$

 $S \sim 2 \log N_c$ Time of first coincidence

Can make estimates for square-root-fewer samples! Can this be extended to nonuniform cases?

- Assumptions needed (won't work always)
- Estimate entropies without estimating distributions.



Slide 9



Generalizing Ma: What is unknown?

Binomial distribution:







Slide 10

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

1-p

р



What is unknown?





Slide 11



For large K the problem is extreme (S known a priori)

 $P_{\beta}(\{q_i\}) = \frac{1}{Z(\beta)} \,\delta\Big(1 - \sum_{i=1}^{K} q_i\Big) \prod_{i=1}^{n} q_i^{\beta-1}$ 0.8 $\beta = 0.0007$, S = 1.05 bits σ Dirichlet priors, a.k.a., (adding pseudocounts 0.2 $\beta = 0.02, S = 5.16$ bits (include the uniform σ prior, the ML prior, and others). 0.01 $\beta = 1, S = 9.35$ bits Inference is analytic σ





Operated by the Los Alamos National Security, LLC for the DOE/NNSA



For large *K* the problem is extreme (S known a priori)

$$P_{\beta}(\{q_i\}) = \frac{1}{Z(\beta)} \,\delta\Big(1 - \sum_{i=1}^{K} q_i\Big) \prod_{i=1}^{K} q_i^{\beta-1}$$

Dirichlet priors, a.k.a., adding pseudocounts (include the uniform prior, the ML prior, and others).

Inference is analytic.





Operated by the Los Alamos National Security, LLC for the DOE/NNSA



Uniformize on S

$$P_{\beta}(\{q_i\},\beta) = \frac{1}{Z} \left. \delta \left(1 - \sum_{i=1}^{K} q_i \right) \prod_{i=1}^{K} q_i^{\beta} \left. \frac{dS}{d\beta} \right|_{N=0} P(S|_{N=0}) \right.$$

- A delta-function sliding along the a priori entropy expectation.
- This is also Bayesian model selection (small β large phase space)
- Have error bars (dominated by posterior variance in β , not at fixed β).



Operated by the Los Alamos National Security, LLC for the DOE/NNSA



Coincidence counting

$$\Delta \equiv N - K_1; \quad K_1 = \text{\#bins with } n_i \ge 1$$

$$\overline{S} = f(\Delta) + \text{corrections}$$

$$\operatorname{var} S = \frac{1}{\Delta} + \text{corrections}$$







Synthetic test (same for natural data)

Refractory Poisson, rate 0.26 spikes/ms, refractory period 1.8 ms, T=15ms, discretization 0.5ms, true entropy 13.57 bits.





NSB summary

- Posterior variance scales as $1/\Delta$
- No bias for short-tailed distributions
- Negative bias for long-tailed distributions (strictly smaller than naïve; as for all learning, cf. Zador and DeWeese)
- Counts coincidences and works in Ma regime (if works)
- Is guaranteed correct (correct) for large *N*
- Smooth convergence: if agrees with itself for different *N*, then correct
- Allows infinite # of bins
- Not a 1/*N* series correction

(Nemenman et al. 2002-2007)



Slide 17



However: recall the slide Exploring the code: long tails



Operated by the Los Alamos National Security, LLC for the DOE/NNSA

MSX

And this slide: Information rate at *T*=25ms



EST 1942

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

- Rate grows up to τ =0.2-0.3 ms
- 30% more information at τ<1ms.
- ~1 bit/spike at 150 spikes/s and lowentropy correlated stimulus. Design principle?
- 0.2 ms comparable to channel opening/ closing noise and experimental noise.



How to estimate entropy for long tails? Go to the source!





Slide 20



What's going on?

- Capture-recapture, but count perches separately from wrasses
- Within each subset, probabilities more uniform
- This is a good convergence test
- But no free lunch: more data needed, since now need coincidences in *each domain*
- Worst case: $2^{(S/2)}$ domains, then $N \sim 2^{S}$
- Usually: N~2^(cS), c<1
- Conjecture: this achieves best possible performance for any distribution, whatever that performance is



Operated by the Los Alamos National Security, LLC for the DOE/NNSA



Example: non NS for generality









How to estimate MI directly?

$I = S_1 + S_2 - S_{12} = (large+large) - (twice large)$ = small, lost in noise



Slide 23



Also recall the slide: Precision is limited by physical noise sources



T = 6 ms $\tau = 0.2 \text{ ms}$ $1.1 \cdot 10^6 \text{ ph/(s \cdot \text{rec})} \pm 3\%$ $I^+ - I^- = 0.0204 \pm 0.0108 \text{ bits}$ p = 6% (and much smaller)

Slide 24



Solution: MI differences (ranks) Non NS, for generality



Slide 25

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

NATIO

EST 1943



Is estimating MI directly worth it? NSB assumtions about MI



Slide 26

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

NATIC

