Randomness and information in biological data
BIO-369

Prof. Anne-Florence Bitbol

EPFL

Lecture 6

Outline of the course

II Extracting information from biological data

1 Quantifying randomness and information in data: entropy
- 1.1 Notion of entropy
- 1.2 Interpretation of entropy
- 1.3 Entropy in neuroscience data: response of a neuron to a sensory input

2 Quantifying statistical dependence
- 2.1 Covariance and correlation
- 2.2 Mutual information
- 2.3 Identifying coevolving sites in interacting proteins using sequence data

3 Inferring probability distributions from data
- 3.1 Model selection and parameter estimation: maximum likelihood
- 3.2 Introduction to maximum entropy inference
- 3.3 Predicting protein structure from sequence data

4 Finding relevant dimensions in data: dimension reduction
- 4.1 Principal component analysis
- 4.2 Beyond principal component analysis

5 Introduction to Bayesian inference

Motivation

Questions:

- "How random" is a random variable?

- How much information are we missing when we don't know the outcome of a random variable (but know its distribution)?

- How much information do we gain when learning about the outcome of a random variable (if we know its distribution)?

→ Can we *quantify* randomness and information?

And also...

- How different are two probability distributions?

- Can we quantify statistical dependence between two random variables?

Do you expect the information content of observing x to be:

A. Independent of P(x)
B. Larger if P(x) is larger
C. Larger if P(x) is smaller
D. It depends

To answer, please:

- Connect to http://ttpoll.eu
- Enter the session ID **bio369**
- Select your answer

Consider a state x that can never be observed, i.e. $P(x)=0$. What do you expect its contribution to entropy to be?

A. 0
B. Infinity
C. 1
D. It depends on the values of $P(x')$ for the other states x'

To answer, please:

- Connect to http://ttpoll.eu
- Enter the session ID **bio369**
- Select your answer

Consider a Bernoulli random variable: P(0)=p and P(1)=1-p. For what value of p do you expect the entropy of this random variable to be maximal?

0% A. For p=0

0% B. For p=1/4

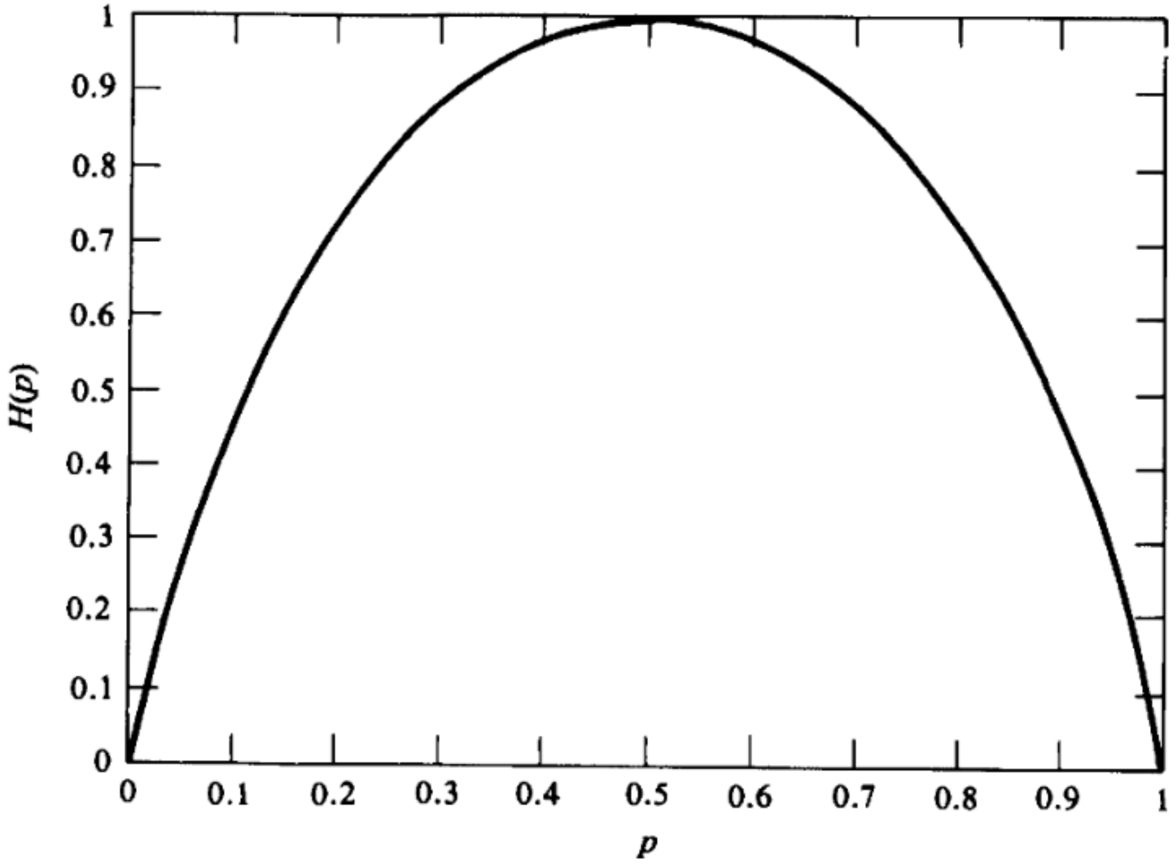
0% C. For p=1/2

0% D. For p=1

To answer, please:

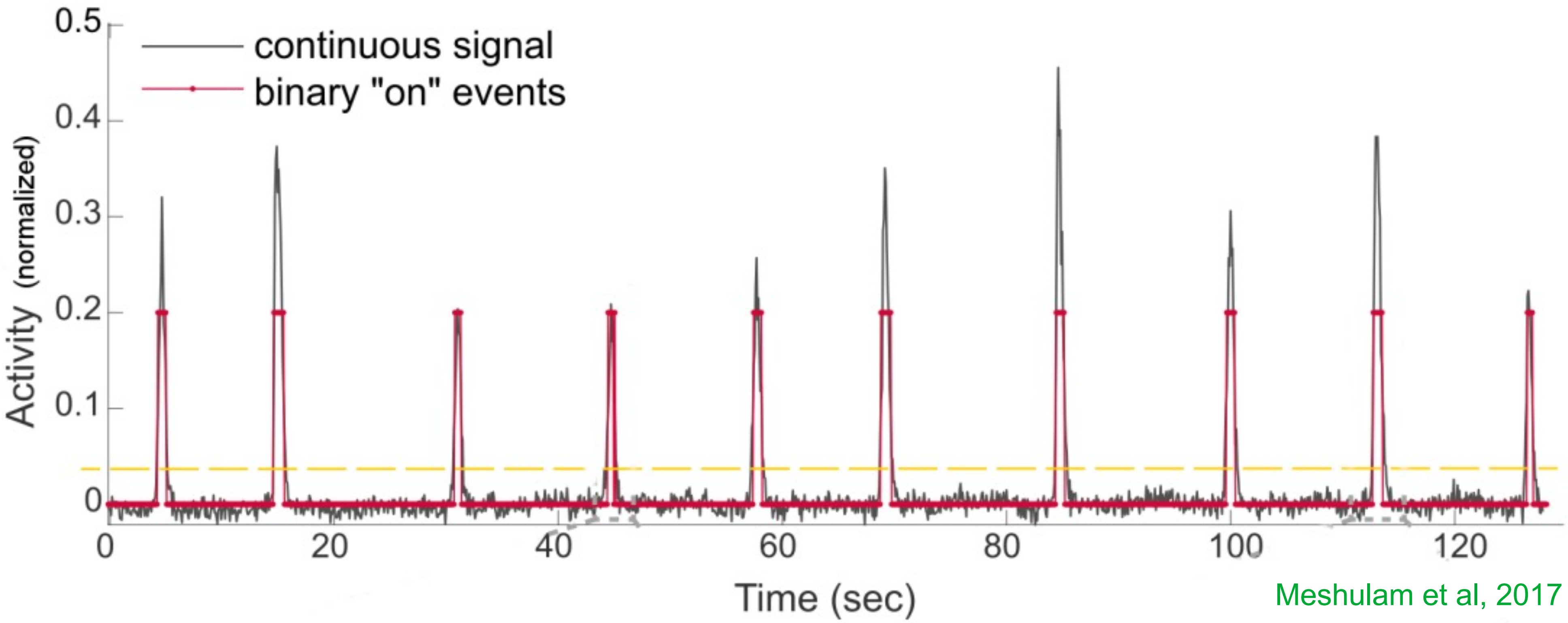
- Connect to http://ttpoll.eu
- Enter the session ID **bio369**
- Select your answer

Entropy of a Bernoulli distribution



Entropy H(p) of a biased coin flip versus probability p of getting “heads” at one flip

Neuroscience data

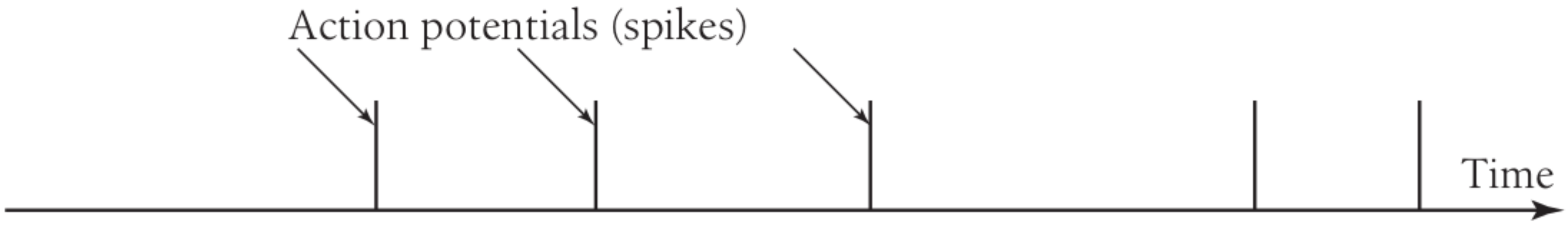


Meshulam et al, 2017

Raw signal from a neuron in units of normalized fluorescence
The series of neuron action potentials is converted to binary on/off activity levels

How does this signal convey information from the stimulus seen by the animal?

From neuroscience data to binary words



- Duration of a spike ~ 0.5 ms
- Minimum time between spikes ~ 3 ms
- Construct bins of duration $\Delta\tau$ in which you count the number of spikes

→ How should we choose the time $\Delta\tau$ in order to get binary (0 or 1) counts?