Signal & Data Analysis in Neuroscience 2020
Information Theory

Izhar Bar-Gad
Room: 408 Phone: 7141 Email: izhar.bar-gad@biu.ac.il

1

Outline

- Entropy
- Mutual information
- Information transmission
- Continuous variables
- Neurons & Entropy

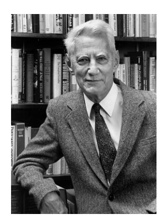
- *Elements of Information Theory*, T. Cover & J. Thomas, Ch. 2.
- *Information Theory, Inference, and Learning Algorithms*, David J.C. MacKay, Ch. 2 (Online version is available on the course web site).

IBG

2

Introduction

- Information theory is a branch of mathematics founded by **Claude Shannon** in the 1940s.
- Information theory sets up **quantitative measures** of information and of the capacity of various systems to transmit, store, and otherwise process information.
- Usage: communication, compression, cryptography, computer science, biology, psychology, neuroscience, etc.



IBG

3

Entropy

- The **entropy** of a system is the amount of ***uncertainty*** about the state of that system.
- The entropy is measured by the number of bits required to fully describe the state of the system.
- Other symbols may easily be transformed to bits e.g. English letters may be represented by 5 bits.
- Could also be thought of as the number of yes/no questions required to establish full understanding.

This type of entropy is also termed Shanon's entropy or Information entropy to distinguish it from the entropy used in Thermodynamics

4

Simple example: coin flipping I

- A coin flip results in either heads or tails. We can mark the outcomes using 1 bit:

  Head = 0 **T**ail = 1

- Following this encoding scheme, the following sequences of coin flips are equivalent:

  H,H,T,H,T ←→ 00101

- Exactly 1 bit is required to represent each toss.

5

Simple example: coin flipping II

- Assuming that we flip two coins simultaneously, we can encode the outcomes as:

Coin A	H	H	T	T
Coin B	H	T	H	T
Encoding	00	01	10	11

- Following this encoding scheme the following sequences of coin flips are equivalent:

00101110 ←→

Trial	1	2	3	4
Coin A	H	T	T	T
Coin B	H	H	T	H

- Exactly 2 bits are required to represent each toss.

6

Simple example: coin flipping III

- What happens if we don't care about the order? We only care if we got both heads, both tails, or a mixed pair.
- The probability of each of these outcomes:
  - both heads - 25%
  - both tails - 25%
  - mixed - 50%
- We will use the following encoding scheme:
  - mixed - 0
  - both heads - 10
  - both tails - 11

7

Simple example: coin flipping IV

- Following this encoding scheme the following sequences of coin flips may be encoded as:

100110 ←

Trial	1	2	3	4
Coin A	H	T	T	T
Coin B	H	H	T	H

- The average number of bits we use:

Both heads:	0.25 x 2 bits	= 0.5 bits
Both tails:	0.25 x 2 bits	= 0.5 bits
Mixes:	0.5 x 1 bit	= <u>0.5 bits</u>
		1.5 bits

8

Entropy & Information

- The **entropy** of a system is the ***uncertainty*** about its state, i.e. the expected number of bits required to fully describe the state of the system.
  - In the final two-coin-flip example, we had a 1.5 bit uncertainty about the outcome.
- **Information** is the amount our uncertainty is reduced given ***new knowledge***.
  - In the two-coin-flip example, if we got new knowledge that the two coins flipped were the same, we will gain 0.5 bits of information (as there is only 1 bit of uncertainty left).

9

Entropy

- Entropy is the expected length in bits of a binary message conveying information
- Other common terms: code complexity, uncertainty, missing/required information, expected surprise, information content, etc.
- Historically, entropy was defined in classic thermodynamics as the "amount of un-usable heat in system" and in statistical thermo-dynamics as the "measure of the disorder in the system", the two were proven to be equivalent.

IBG

10

Shannon Information

- Smallest unit of information is the "bit"
- 1 bit = the amount of information needed to choose between two equally-likely outcomes (e.g. flip a coin)
- Properties:
  - Information for independent events adds
  - Information is zero if we already know the outcome

IBG

11

Shannon Information: Surprise I

The surprise of a single event is high for unexpected (low probability) events and low for expected events.

$$p(r_1) = 1 \quad \Rightarrow \quad h(p(r_1)) = 0$$

$$p(r_2) \to 0 \quad \Rightarrow \quad h(p(r_2)) \to \infty$$

Independent events: $p(r_1, r_2) = p(r_1)\,p(r_2)$

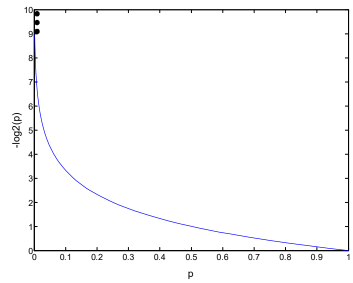
Implies: $h(p(r_1, r_2)) = h(p(r_1)) + h(p(r_2))$

$$\boxed{h(p(r)) = -\log_2(p(r))}$$

IBG

12

Shannon Information: Surprise II



13

Logarithms – useful formulas

$$\log_a X \cdot Y = \log_a X + \log_a Y$$

$$\log_a \frac{X}{Y} = \log_a X - \log_a Y$$

$$\log_a X^Y = Y \log_a X$$

$$\log_a X = \frac{\log_b X}{\log_b a}$$

$$\frac{d \log_a X}{dX} = \frac{\log_a e}{X}$$

14

Entropy - definition

- Entropy is the mean value of the surprise over all possible observations

$$H(X) = E_p[-\log_2 p(x)]$$

- In the discrete case:

$$H(X) = -\sum_x p(x) \log_2 p(x)$$

15

Example: a two outcome event I

- The entropy of the result of a fair coin toss:

$$H = -[0.5 \cdot \log_2(0.5) + (1-0.5) \cdot \log_2(1-0.5)]$$
$$= -[-0.5 - 0.5] = 1$$

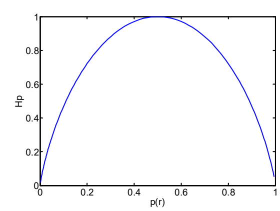
- The entropy of an unfair (99% head) coin toss:

$$H = -[0.99 \cdot \log_2(0.99) + (1-0.99) \cdot \log_2(1-0.99)]$$
$$= -[-0.0144 - 0.0644] = 0.08$$

16

Example: a two outcome event II

- In the general case:



$$H = -[p \cdot \log_2(p) + (1-p) \cdot \log_2(1-p)]$$

17

Entropy properties

- Entropy is always positive
- Entropy is maximum if p(r) is constant
- Entropy is minimum if p(r) is a delta function
- The higher the entropy, the more you learn (on average) by observing values of the random variable
- The higher the entropy, the less you can predict the values of the random variable

18

Calculating Entropy: The simple case

- If all n possible outcomes of situation X are equally probable, then our uncertainty about which one will occur can be calculated by:

$$H(X) = \log_2(n) \text{ bits}$$

- Out of gold eight coins, one of which is a fake, while you know the other seven are real. You know the fake one has a different weight than the rest. How many weightings on a balance scale will it take to determine the fake? What if you only had seven coins with one fake? What if you had nine coins with one fake?

IBG

19

Encoding based on entropy I

- Suppose we have 4 symbols: A C G T with
- The symbol probabilities are:
  $P_a = 0.5$ $P_c = 0.25$ $P_g = P_t = 0.125$
- Leading to surprises:
  $h_a = 1$bit $h_c = 2$bit $h_g = h_t = 3$ bit
- Thus the mean uncertainty of a symbol is:
  H = 1*0.5+2*0.25+0.125*3+0.125*3 = 1.75 bit

IBG

20

Encoding based on entropy II

- One option for encoding uses 2 bits for each symbol: A = 00 C = 01 G = 10 T = 11
- In the other option the number of binary digits equals the surprise: A = 1 C=01 G=000 T=001
- So the string ***ACATGAAC*** which has frequencies the same as the probabilities defined above, is coded as:

Method 1	0001001110000001	16 (2 bits per symbol)
Method 2	10110010001101	14 (1.75 bits per symbol)

IBG

21

Encoding based on entropy III

- In this specific case, can we find a better (shorter) encoding ?
- In the general case, how can we formulate the optimal encoding ?
- These questions are handled under the data compression topic…

Elements of Information Theory, T. Cover & J. Thomas, Chapter 5.

IBG

22

Outline

- Entropy
- Mutual information
- Information transmission
- Continuous variables
- Neurons & Entropy

IBG

23

Joint entropy

- The joint entropy may be considered a single vector valued random variable:

$$H(X,Y) = E_{p(x,y)}[-\log_2 p(x,y)]$$

- In the discrete case:

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

IBG

24

Conditional entropy

Same formulation, but using the conditional density:

$$H(Y|X) \stackrel{\text{def}}{=} \sum_{x\in\mathcal{X}} p(x)\, H(Y|X=x)$$
$$= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x)\ \log\ p(y|x)$$
$$= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(y,x)\ \log\ p(y|x)$$
$$= -E_{p(x,y)} \log\ p(y|x).$$

25

The conditional entropy chain rule

$$H(Y|X) = H(Y,X) - H(X)$$

Proof:

$$H(Y|X) = -E_{p(x,y)} \log\, p(y|x)$$
$$= -E_{p(x,y)} \log \left(\frac{p(y,x)}{p(x)}\right)$$
$$= -E_{p(x,y)}(\log p(y,x) - \log p(x))$$
$$= -E_{p(x,y)} \log p(y,x) + E_{p(x)} \log p(x)$$
$$= H(Y,X) - H(X).$$

Thus:

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$$
$$H(Y \mid X) = H(X \mid Y) + H(Y) - H(X)$$

26

Mutual information I

- The entropy tells us how much we can learn (therefore how much we don't know)
- The mutual information between *r* and *s* is:
  - How much do we learn about *r* by observing *s?*
  - How much more do we know about *r* after observing *s*?
  - How much easier is it to predict *r* after observing *s*?
- Therefore: How much has the entropy of *r* decreased after observing *s*?

27

Mutual information II

- Mutual information = How is the entropy of *r* decreased by knowing *s*?

$$H(R|S) = -\sum_r \sum_s p(r,s) \cdot \log(p(r,s))$$

$$\begin{aligned} I(R;S) &= H(R) - H(R\,|\,S) \\ &= -\sum_r p(r)\cdot\log(p(r)) + \sum_r\sum_s p(r,s)\cdot\log(p(r\,|\,s)) \\ &= -\sum_r\sum_s p(r,s)\cdot\log(p(r)) + \sum_r\sum_s p(r,s)\cdot\log(p(r\,|\,s)) \\ &= \sum_r\sum_s p(r,s)\cdot\left[-\log(p(r)) + \log(p(r\,|\,s))\right] \\ &= \sum_r\sum_s p(r,s)\cdot\log\left(\frac{p(r\,|\,s)}{p(r)}\right) = \sum_r\sum_s p(r,s)\cdot\log\left(\frac{p(r,s)}{p(r)\cdot p(s)}\right) \end{aligned}$$

28

The doctor example I

- We're back to the doctor who need to distinguish between:
  - The flu $p(x_1)$=0.9
  - Severe infection $p(x_2)$ =0.1
- He has two tests:

Blood test Y	Flu	Infection
Positive	0.2	0.7
Negative	0.8	0.3

Urine test Z	Flu	Infection
Positive	0.1	0.5
Negative	0.9	0.5

- **Which test gives more information about the state of the patient?**

29

The doctor example II

$$I_m = \sum_{s,r} P[s]\,P[r|s]\,\log_2\left(\frac{P[r|s]}{P[r]}\right)$$

$P(y_+)$=0.9*0.2+0.1*0.7=0.25 $P(y_-)$ = 0.75
$P(z_+)$=0.9*0.1+0.1*0.5=0.14 $P(z_-)$ = 0.86

H(X)=-(0.9*$\log_2$(0.9)+0.1*$\log_2$(0.1)) = 0.436

I(Y;X)=0.9*0.2*$\log_2$(0.2/0.25)+0.9*0.8*$\log_2$(0.8/0.75)+ 0.1*0.7*$\log_2$(0.7/0.25)+0.1*0.3*$\log_2$(0.3/0.75) = 0.0734

I(Z;X)=0.9*0.1*log2(0.1/0.14)+0.9*0.9*$\log_2$(0.9/0.86)+ 0.1*0.5*$\log_2$(0.5/0.14)+0.1*0.5*$\log_2$(0.5/0.86) = 0.0621

Thus, the blood test is more informative…

30

Properties of mutual information I

- Zero if r and s are independent
  $p(r,s) = p(r)p(s) \Rightarrow I(R,S) = 0$
- Cannot be more than the entropy
  $I(R,S) \leq H(R) \quad I(R,S) \leq H(S)$
- Cannot be increased by math alone
  $I(f(R),S) \leq I(R,S)$

Holds true for any function, so no transmission line, neural network, or computation can ever squeeze out more information!

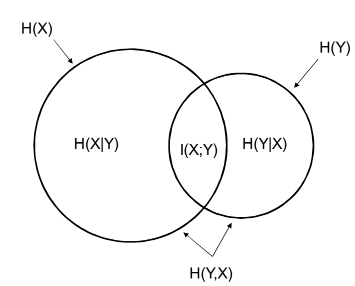
31

Properties of mutual information II

- I(X;Y)=H(X)-H(X|Y)
- I(X;Y)=H(Y)-H(Y|X)
- I(X;Y)=H(X)+H(Y)-H(Y,X)
- I(X;Y)= I(Y;X)
- I(X;X)=H(X)

32

Entropy and Mutual information



33

Relative entropy ≡ Kullback Liebler (KL) divergence

The Kullback-Leibler (KL) divergence is a 'distance' measure between probability distributions.

$$D_{KL}(p,q) = \sum_r p(r) \log_2 \frac{p(r)}{q(r)}$$

$D_{KL}(p,q) \neq D_{KL}(q,p)$, and $D_{KL} \geq 0$

Thus,

$$I_m = D_{KL}(p(r,s), p(r)p(s))$$

- The excess message length needed to use p(x) - optimized code for messages based on q(x)

34

Relative entropy properties

$$I_m = \sum_{s,r} P[r,s] \log_2 \left(\frac{P[r,s]}{P[r]P[s]} \right)$$

$$D_{KL}(p,q) = \sum_r p(r) \log_2 \frac{p(r)}{q(r)}$$

$$D_{KL}(p(r,s), p(r)p(s)) = \sum_{r,s} p(r,s) \log_2 \frac{p(r,s)}{p(r)p(s)}$$

$$D_{KL}(p(r)p(s), p(r,s)) = \sum_{r,s} p(r)p(s) \log_2 \frac{p(r)p(s)}{p(r,s)}$$

$$\downarrow$$

$$I_m = D_{KL}(p(r,s), p(r)p(s))$$
$$I_m = D_{KL}(p(s,r), p(s)p(r))$$
$$I_m \neq D_{KL}(p(r)p(s), p(r,s))$$

35

Additional (in) equalities

- $D_{KL}(p,q) \geq 0$ (information inequality)
  $D_{KL}(p,q)=0$ iff p(x)=q(x) for every x
- I(X;Y)≥0 (Non negativity of mutual information)
  I(X;Y)=0 iff Y & X are independent
- H(X|Y)≤H(X) (Conditioning reduces entropy)
- $H(X_1, X_2, ..., X_n) \leq \sum_{i=1}^{n} H(X_i)$ (Independence bound)

If *f* is convex → *E(f(X))≥f(E(X))* (Jensen inequality)

36

Outline

- Entropy
- Mutual information
- Information transmission
- Continuous variables
- Neurons & Entropy

Elements of Information Theory, T. Cover & J. Thomas, Chapter 8.

IBG

37

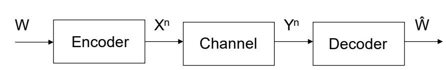
Compression vs. Transmission

- During compression all the redundancy is removed from the data.
- During transmission redundancy is added to the data to enable error correction.

IBG

38

Information transmission



- **Discrete channel** – transitioning between alphabet X to Y through a probability matrix p(y|x).
- **Memoryless channel** – the probability distribution of Y depends only on the input at the same time.

- The challenge is encoding the message in such a way that it occupies minimal space while still containing enough redundancy to be able to detect and correct errors.

IBG

39

Channel examples

1. A word *W* in English may be transformed into a series of syllables via speech which are passed through the air channel and upon hearing converted back to a series of syllables and to the reconstructed word.
2. A word *W* in English may be transformed into a series of letters represented by 8 bit ASCII code and passed through a communication line and upon receiving at a different computer transformed back to a series of letters and to the reconstructed word.

IBG

40

Properties of Channels

- Each channel has a transmission rate – the number of symbols it can transmit per time unit.
- Channels have error rates, which determine, for any particular symbol, the probability that a different symbol will come out of the channel.
- The error rate of the channel determines its capacity - the bits of *information* that are transmitted per symbol sent.
- The transmission rate and the channel capacity can be multiplied to get its data rate - the rate at which information can be sent across the channel.

IBG

41

Channel capacity

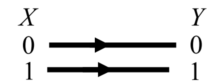
For a discrete memoryless channel: the capacity is limiting information transport rate that can be achieved with vanishingly small error probability.

$$C = \max_{p(x)} I(X;Y)$$

IBG

42

Noiseless binary channel



- Assuming a binary alphabet for both X & Y and a noiseless channel:

$$I(X;Y) = H(X) - H(X \mid Y) = H(X) - 0 = H(X)$$

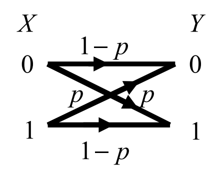
- The channel capacity is maximal when:

$$p(x=1) = p(x=0) = 0.5 \quad \Rightarrow \quad C = 1$$

43

Noisy binary symmetric channel

$$X = \{0,1\} \quad Y = \{0,1\}$$



$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

44

Binary symmetric channel – Mutual information II

$$\begin{aligned} I(X;Y) = \quad & H(Y) - H(Y|X) \\ = \quad & H(Y) - \sum p(x) H(Y|X=x) \\ = \quad & H(Y) - \sum p(x) H(p) \\ = \quad & H(Y) - H(p) \\ \leq \quad & 1 - H(p) \end{aligned}$$

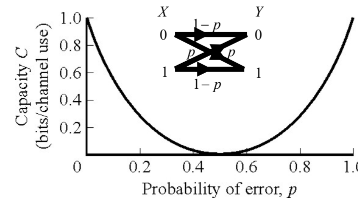
- Equality is achieved:
  - P(y=1)=P(y=0)=0.5 → P(x=1)=P(x=0)=0.5

$$C = 1 - H(p)$$

45

Binary symmetric channel – Channel capacity

When p=1 bits are inverted but information is perfect if invert them back!



$$C = (1-p)\log_2 2(1-p) + p\log_2 2p$$

46

Example - calculating capacity

- Inputs & Outputs are binary.
- Maximal input uncertainty: H(Input) = 1
- Given a 1% error rate:

H(Input|Output) = - (0.99 log 0.99 + 0.01 log 0.01)
= 0.0144 + 0.0664 = **0.0808** bits
I(Input;Output) = H(Input) - H(Input | Output)
= 1 - 0.0808 = **0.9192** bits

- This is also the capacity since it is the maximal input/output information.

47

Dealing with Errors…

- Assuming we know that there are going to be some errors, how can we be sure to get our information across?
- If we're really unlucky, we can't. But we can make sure to be able to tolerate any reasonable amount of error.
- What's one way for us to be able to be sure we can detect any single error in our message?
- How can we make sure we can *correct* any error in the message?

48

How good is Error Correction?

• We can do better. We can get as close to the channel capacity as we want, though we may need long messages.

• The channel capacity is defined as the information that passes through the channel.

• If we are correct in our definition of information, it should give us a perfect measure of how many bits we can send through the channel.

• Intuitively channel capacity makes sense. We start with maximal uncertainty about the symbol that entered the channel. That uncertainty is lowered when we see a symbol come out.

IBG

49

Channel coding theorem

- An (M, n) code is:
  - Index set {1,2,…,M}
  - Encoding function
    {1, …, M} → {$X^n(1)$, …, $X^n(M)$}
  - Decoding function
    Y^n → {1, …, M}
- The rate of an (M, n) code is: $R = \frac{\log M}{n}$
- The rate is *achievable* if there exists a sequence (2^{nR}, n) leading to an error→0 for n→∞.
- **All the rates below the channel capacity are achievable (R≤C).**

IBG

50

Outline

- Entropy
- Mutual information
- Information transmission
- Continuous variables
- Neurons & Entropy

Elements of Information Theory, T. Cover & J. Thomas, Chapter 9.

IBG

51

Continuous variables

- A real number has an infinite number of bits, therefore theoretically, **infinite information**.
- However, there is always noise (or quantization) which defines a number of discriminable levels

52

Entropy & Differential entropy

- Usage of probability density instead of probability

$$\begin{aligned} H &= -\sum p[r]\Delta r \log_2 (p[r]\Delta r) \\ &= -\sum p[r]\Delta r \log_2 p[r] - \log_2 \Delta r \end{aligned}$$

- Note: for Δr→0 the log diverges…

$$h(r) = \lim_{\Delta r}\{H(r) + \log_2 \Delta r\} = -\int p(r)\log_2 p(r)dr$$

53

Differential entropy

$$\boxed{h(x) = -\int p(x)\log_2 p(x)dx}$$

- Example 1: **Uniform distribution** (interval [0,a])

$$h(x) = -\int_0^a \frac{1}{a}\log_2 \frac{1}{a}dx = -\log_2 \frac{1}{a} = \log_2 a$$

Note: for a<1 the differential entropy is negative

- Example 2: **Normal distribution** (μ=0,σ)

$$h(x) = \frac{-1}{\sigma\sqrt{2\pi}}\int e^{-x^2/2\sigma^2}\log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\right)dx = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

54

Entropy of a sampled continuous variable

- Following a *n* bit quantization of the variable (i.e. accuracy of 2^{-n})
  $H(X)=h(X)-\log(2^{-n})=h(X)+n$
- Example: a uniform distribution over the interval [0,1] with a resolution of ~0.001
  $H(X)=\log_2(1)+\log_2(1000)\sim 10$
- Example: a uniform distribution over the interval [0,¼] with a resolution of ~0.001
  $H(X)=\log_2(1/4)+\log_2(1000)\sim 8$
  Since the first two bits are always 0.

55

Outline

- Entropy
- Mutual information
- Information transmission
- Continuous variables
- Neurons & Entropy

Theoretical Neuroscience, Peter Dayan & Larry Abbott, Ch. 4.
Spikes, F. Rieke, D. Warland, R. van Steveninck & W. Bialek.

56

Neurophysiological based information theoretic questions

- How much information do the neurons **convey**?
- How much information is conveyed through a **spike**?
- How much does spiking activity tell us about a stimulus?
- Is the neural representation **optimal**?
- Is the information encoded by a neuronal population **redundant**?
- Can **rate** by itself encode all the information?
- Is there and if so, what is the **theoretical limit** on the information in the nervous system?

57

Rate encoding – maximum entropy I

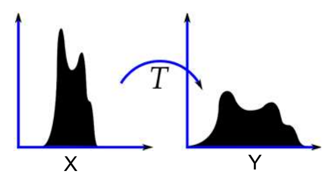
- **If** information is conveyed by the firing rate r, all firing rates should have equal probability.
- For a neuron with a rate $r_{range} = r_{max} - r_{min}$

$$p(r) = \frac{1}{r_{range}}$$

- Thus, when the rate represents another non-uniform variable, maximal entropy will be achieved through **histogram equalization**.

58

Histogram equalization



- What is the transfer function, *T* ?

59

Rate encoding – maximum entropy II

$$\frac{|r(s+\Delta s) - r(s)|}{r_{range}} = p(s) \cdot \Delta s$$

- If *s* is not uniformly distributed, then need to adjust, for monotonically increasing *r(s)*:

$$\frac{dr}{ds} = r_{range} \cdot p(s) \qquad \Rightarrow \qquad \frac{dr}{ds} \propto p(s)$$

- Assign more bits to regions of higher probability.

60

Maximum entropy for a population

- For a population maximum, every neuron must have maximum entropy by itself.
- Two neurons firing with identical mean rates are the same as one neuron firing for twice as long leading to an entropy which is proportional to the number of neurons.

$$H_{r_1,r_2} \le H_{r_1} + H_{r_2} \quad (= iff\ r_1\ and\ r_2\ independent)$$

- This type of independent coding is usually termed "Factorial code".

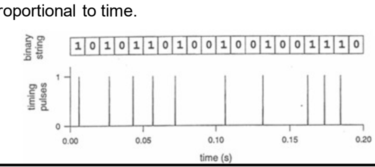
61

Entropy of a spike train

- How many different patterns can occur over a fixed length T?
- If all bins are independent, this is equivalent to tossing a biased coin T/Δt times.
- Each toss has: $H = -((r\Delta t)\log_2(r\Delta t) + (1 - r\Delta t)\log_2(1 - r\Delta t))$

$$H_{total} = -\frac{T}{\Delta t}\cdot(r\cdot\Delta t\cdot\log_2(r\cdot\Delta t) + (1 - r\cdot\Delta t)\cdot\log_2(1 - r\cdot\Delta t))$$

Proportional to time.



62

Information for spike trains

- Need to consider every pattern of spikes over an interval T as being a single binary number.
- Many possible binary numbers; may be difficult to estimate p(r|s) unless T is very short.
- Information rate is the **bits per second** (or **bits per spike**) related to the input
- If the chance of a spike in a bin is small (low rate, or high sampling rate) then we can approximate the entropy rate as:

$$H/T \approx -r\log_2(r\Delta t)$$

63

Encoding – spike time vs. count

- What is the maximal information using a spike count measure vs. the spike timing?
- Example: assuming a neuron with 3ms refractory period what is the maximal entropy given 10 successive bins of 3ms each holding a maximum of one spike of vs. one bin of 30ms allowing a maximum of 10 spikes?
- Is the neuron conveying information when it is not firing?

IBG

64

Neurophysiological based information theoretic questions

- How much information do the neurons **convey**?
- How much information is conveyed through a **spike**?
- How much does spiking activity tell us about a stimulus?
- Is the neural representation **optimal**?
- Is the information encoded by a neuronal population **redundant**?
- Can **rate** by itself encode all the information?
- Is there and if so, what is the **theoretical limit** on the information in the nervous system?

IBG

65