THEORETICAL NEUROSCIENCE I

Lecture 17: Conditional entropy and mutual information

Prof. Jochen Braun

Otto-von-Guericke-Universität Magdeburg,
Cognitive Biology Group

Outline

1. The story so far . . .

2. Quantifying shared entropy or information

3. Bivalued stimulus and response

4. Information of spiking neurons

5. Neurons with variable firing rates (advanced)

6. Histogram equalization (advanced)

1 The story so far ...

1. We have considered ensembles X of discrete outcomes x, each with a probability $P(x)$.

The information content of each outcome x can be measured in *bits*. It is the minimal number of *binary questions* needed to determine an outcome.

The average information content over all possible outcomes is the *Shannon entropy*:

$$H(X) = \sum_x P(x) \log_2 \frac{1}{P(x)}$$

Each individual outcome x contributes

$$P(x) \log_2 \frac{1}{P(x)}$$

- The information content of an event depends solely on its probability.
- Not nature of events, but number and frequency of alternative events is what matters.
- Syntax, not semantics!
- Focus shifts from what happened, to how often it happens and to what else could have happened, but did not!
- Scope of enquiry is enlarged to previously ignored facts and phenomena!

- Statistics of neuronal and sensory events becomes important (e.g., in natural environments).

Joint ensembles.
Further, we wished to understand information sharing between two ensembles X and Y, each with distinct outcomes x and y.

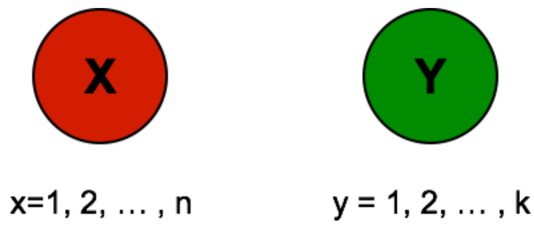

Figure 1: Joint ensembles.

Joint probability
Jointly observing both ensembles, we considered the probabilities of joint outcomes x, y:

	$x = 1$	$x = 2$	$\dots$
$y = 1$	$P(1, 1)$	$P(1, 2)$	$\dots$
$y = 2$	$P(2, 1)$	$P(2, 2)$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\ddots$

Joint entropy and individual entropies

Given the joint probabilities, we defined the *joint entropy* of X and Y as

$$H(X, Y) = \sum_x \sum_y P(x, y) \log_2 \frac{1}{P(x, y)}$$

It compares to the *individual* or *marginal* entropies of X, and of Y,

$$H(X) = \sum_x P(x) \log_2 \frac{1}{P(x)}, \qquad H(Y) = \sum_y P(y) \log_2 \frac{1}{P(y)}$$

Joint entropy *cannot be larger* than the sum of individual entropies

$$H(X, Y) \leq H(X) + H(Y)$$

2 Quantifying shared entropy or information

To better quantify the sharing of entropy or information, we now introduce two further notions

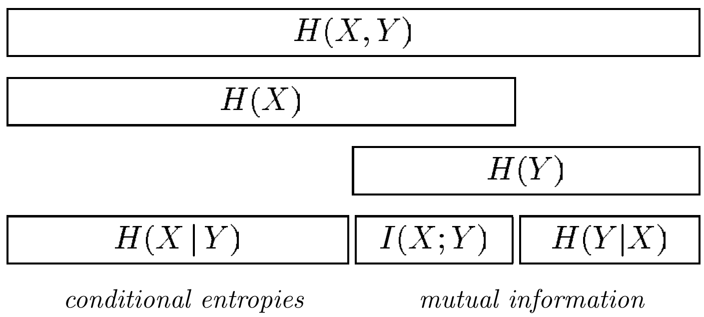

Figure 2: Quantifying shared entropy or information.

Conditional entropy, noise entropy

The **conditional probability** $P(y|x)$ is defined for a subset of outcomes (x, y): it is the probability to find outcome y with in the subset of outcomes x.

The **conditional entropy** is the entropy of the conditional probability. It reveals how much variability remains in y, when x is fixed. By definition, this part of the entropy of y is uninformative about x (which is given). Therefore, this part of the entropy of y is *not shared* with x.

Because the conditional entropy of y is uninformative about x, it is also called the **noise entropy**.

The conditional/noise entropy of Y, given $X = x_0$

... is the entropy of the probability distribution $P(y|x = x_0)$

$$H(Y|X = x_0) = \sum_y P(y|x = x_0) \log_2 \frac{1}{P(y|x = x_0)}$$

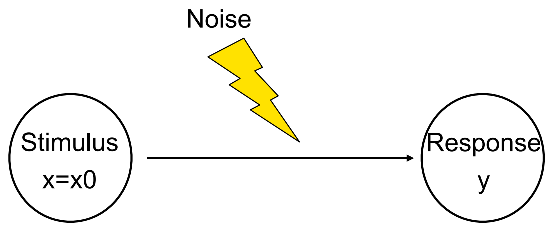

Figure 3: The conditional/noise entropy of Y, given $X = x_0$.

The conditional/noise entropy of Y, given X.

... is the average, over all possible x_0, of the conditional entropy of Y, given $X = x_0$:

$$H(Y|X) = \sum_{x_0} P(x_0) \left[\sum_y P(y|x = x_0) \log_2 \frac{1}{P(y|x = x_0)} \right] =$$

$$= \sum_{xy} P(y|x) P(x) \log_2 \frac{1}{P(y|x)} = \sum_{xy} P(x, y) \log_2 \frac{1}{P(y|x)}$$

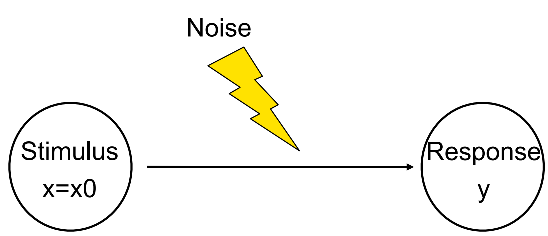

Figure 4: The conditional/noise entropy of Y, given X.

The conditional/noise entropy of X, given $Y = y_0$

... is the entropy of the probability distribution $P(x|y = y_0)$

$$H(X|y = y_0) = \sum_x P(x|y = y_0) \log_2 \frac{1}{P(x|y = y_0)}$$

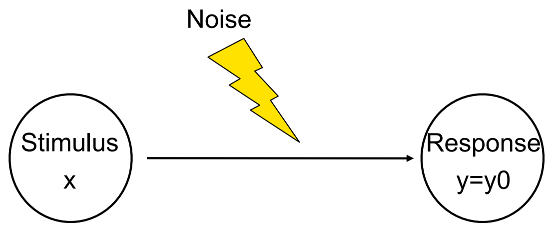

Figure 5: The conditional/noise entropy of X, given $Y = y_0$.

The conditional/noise entropy of X, given Y

... is the average, over y, of the conditional entropy of X, given y:

$$H(X|Y) = \sum_{y_0} P(y_0) \left[\sum_x P(x|y = y_0) \log_2 \frac{1}{P(x|y = y_0)} \right]$$

$$= \sum_{xy} P(x|y) P(y) \log_2 \frac{1}{P(x|y)} = \sum_{xy} P(x, y) \log_2 \frac{1}{P(x|y)}$$

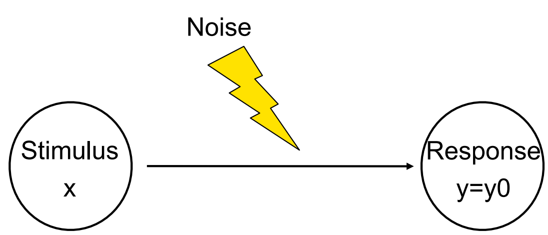

Figure 6: The conditional/noise entropy of X, given Y.

Conditional entropy, noise entropy

- Conditional entropy of A (given B) is the remaining entropy of A when B is fixed.
- Thus, conditional entropy of A (given B) is the part of the entropy of A that is *uninformative* about B.
- For this reason, it is also called the 'noise entropy' of A (with respect to B).
- From additivity, it follows that any entropy of A beyond this 'noise entropy' is *informative* about B.
- Thus, the 'informative' entropy of A is the total entropy *minus* the 'noise' entropy.

Chain rule for information content

From the product rule for probabilities

$$P(x, y) = P(x|y)\, P(y)$$

$$\log_2 \frac{1}{P(x, y)} = \log_2 \frac{1}{P(x|y)} + \log_2 \frac{1}{P(y)}$$

we obtain

$$h(x, y) = h(x|y) + h(y)$$

Thus, the information of x and y is the information of x given y, plus the information of y.

Chain rule for entropy

The joint, conditional, and marginal entropies are related as follows:

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

In words, the joint entropy of X and Y is the conditional entropy of X given Y, plus the marginal entropy of Y (and *vice versa*).

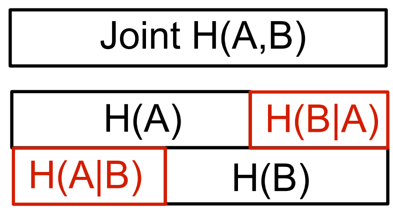

Figure 7: Chain rule for entropy.

Chain rule (optional)

$$\begin{aligned}
H(X|Y) \;+\; H(Y) &= \\
&= \sum_{xy} P(x,y)\,\log_2 \frac{1}{P(x|y)} + \sum_{y} P(y)\log_2 \frac{1}{P(y)} = \\
&= \sum_{xy} P(x,y)\,\log_2 \frac{1}{P(x|y)} + \sum_{xy} P(x,y)\log_2 \frac{1}{P(y)} = \\
&= \sum_{xy} P(x,y)\,\left[\log_2 \frac{1}{P(x|y)} + \log_2 \frac{1}{P(y)}\right] = \\
&= \sum_{xy} P(x,y)\,\log_2 \frac{1}{P(x|y)\,P(y)} = \\
&= \sum_{xy} P(x,y)\,\log_2 \frac{1}{P(x,y)} = H(X,Y)
\end{aligned}$$

The mutual information between X and Y

$$I_m = H(X) - H(X|Y) \geq 0$$

is here defined as

- the information that x conveys about y.
- the total information conveyed by x minus the noise entropy of x.
- the average reduction in uncertainty about y from learning the value of x.

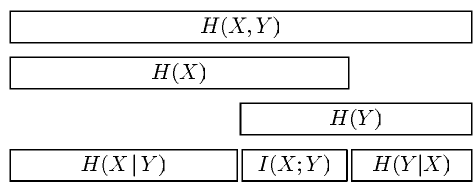

Figure 8: The mutual information 1.

Mutual information is commutative and also satisfies

$$I_m = H(Y) - H(Y|X) \geq 0$$

which is

- the average amount of information that y conveys about x.
- the total information conveyed by y minus the noise entropy of y.
- the average reduction in uncertainty about x from learning the value of y.

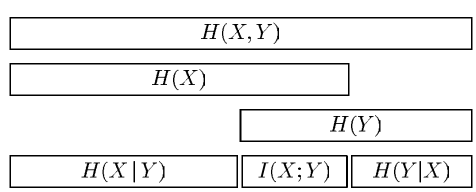

Figure 9: The mutual information 2.

Yet another way to express the mutual information is

$$I_m = H(X) + H(Y) - H(X,Y)$$

which is

- the difference between the *total individual* entropies and the *joint* entropy.

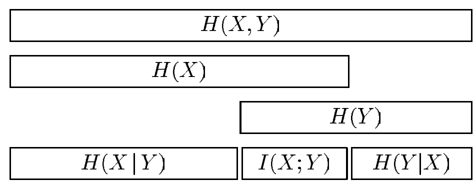

Figure 10: The mutual information 3.

Alternative expressions for MI

The mutual information can be expressed in several ways!

Marginal x minus conditional on y:

$$I_m = H(X) - H(X|Y) = \sum_x P(x)\log_2 \frac{1}{P(x)} - \sum_{xy} P(x,y)\ \log_2 \frac{1}{P(x|y)}$$

Marginal y minus conditional on x:

$$I_m = H(Y) - H(Y|X) = \sum_y P(y)\log_2 \frac{1}{P(y)} - \sum_{xy} P(x,y)\ \log_2 \frac{1}{P(x|y)}$$

Total marginal minus joint:

$$I_m = H(X) + H(Y) - H(X,Y) = \sum_{xy} P(x,y)\ \log_2 \frac{P(x,y)}{P(x)\,P(y)}$$

Mutual information (optional)

$$
\begin{aligned}
H(X) \; - \; H(X|Y) = & \\
= & \sum_x P(x) \log_2 \frac{1}{P(x)} - \sum_{xy} P(x,y) \log_2 \frac{1}{P(x|y)} = \\
= & \sum_{xy} P(x,y) \log_2 \frac{1}{P(x)} - \sum_{xy} P(x,y) \log_2 \frac{1}{P(x|y)} = \\
= & \sum_{xy} P(x,y) \left[\log_2 \frac{1}{P(x)} - \log_2 \frac{1}{P(x|y)} \right] = \\
= & \sum_{xy} P(x,y) \log_2 \frac{P(x|y)\, P(y)}{P(x)\, P(y)} = \\
= & \sum_{xy} P(x,y) \log_2 \frac{P(x,y)}{P(x)\, P(y)} = H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

Summary mutual information

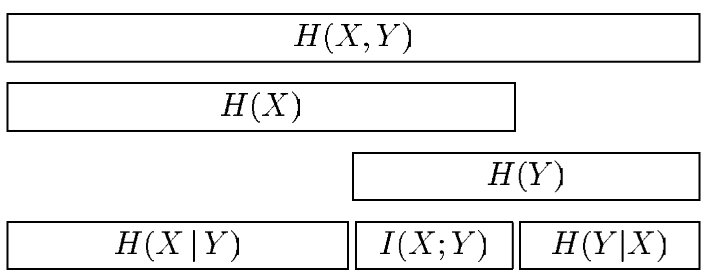

Figure 11: The mutual information 4.

3 Bivalued stimulus and response

In a joint ensemble $P(R, S)$, the value of r is probabilistically related to the value of s:

$$P(s) = \begin{cases} s_+ & s = + \\ 1 - s_+ & s = - \end{cases} \qquad P(r) = \begin{cases} r_+ & r = + \\ 1 - r_+ & r = - \end{cases}$$

$$P(r|s) = \begin{cases} x & s = r \\ 1 - x & s \neq r \end{cases}$$

The coupling parameter $x \in [0, 1]$ determines the probability that r and s take equal values.

For $x = 0$ or $x = 1$, the joint ensemble r, s is *deterministically dependent*; for $0 < x < 1$ it is *probabilistically dependent*.

Joint and individual distribution r (optional)

The joint distribution is the product of $P(r|s)$ and $P(s)$:

$P(r,s)$	s		
		$+$	$-$
r	$+$	$x\, s_+$	$(1-x)\,(1-s_+)$
	$-$	$(1-x)\, s_+$	$x\,(1-s_+)$

and the individual distribution is $P(r) = \sum_s P(r, s)$:

$$P(+) = r_+ \;=\; 1 - s_+\,(1-x) - x\,(1-s_+)$$

$$P(-) = r_- \;=\; s_+\,(1-x) + x\,(1-s_+)$$

Entropies and mutual information (optional)

We are now in the position to compute individual and joint entropies and mutual information:

$$H(S) = s_+ \log_2 \frac{1}{s_+} + (1 - s_+) \log_2 \frac{1}{1 - s_+}$$
$$H(R) = r_+ \log_2 \frac{1}{r_+} + (1 - r_+) \log_2 \frac{1}{1 - r_+}$$

$$H(S,R) = x\, s_+ \log_2 \frac{1}{x s_+} + (1-x)(1-s_+) \log_2 \frac{1}{(1-x)(1-s_+)} + \\ + (1-x)\, s_+ \log_2 \frac{1}{(1-x)\, s_+} + x\,(1-s_+) \log_2 \frac{1}{x\,(1-s_+)}$$

$$I_m = H(S) + H(R) - H(S,R)$$

Stimulus entropy

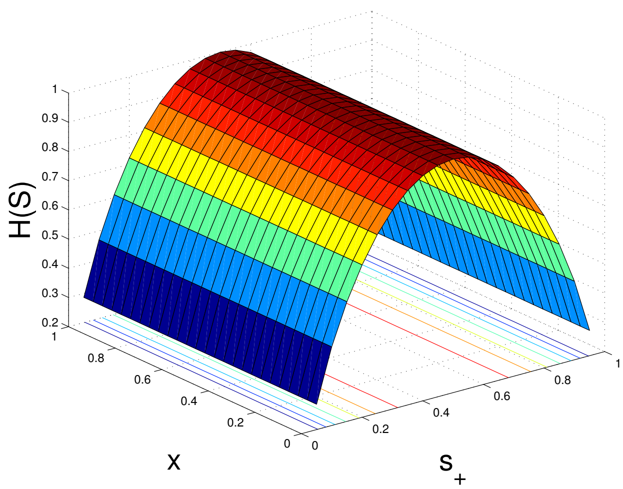

Figure 12: Stimulus entropy.

Stimulus entropy as a function of stimulus probability s_+. Note deterministic ($s_+ = 0\, or\, 1$) and probabilistic regimes ($0 < s_+ < 1$)

Response entropy

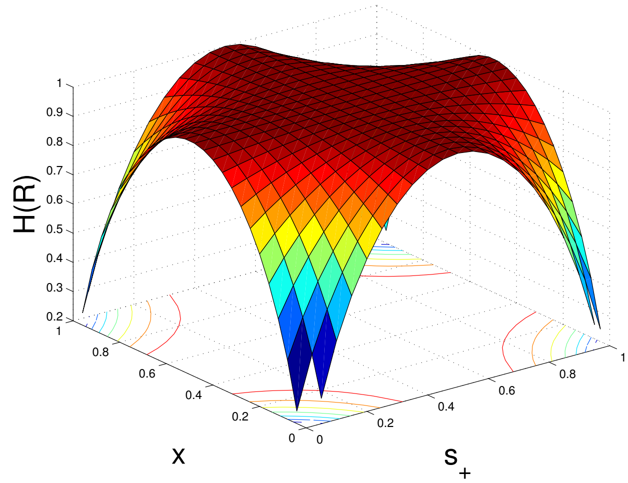

Figure 13: Response entropy.

Response entropy as a function of stimulus probability s_+ and parameter x. Note *additional* probabilistic regimes ($0 < x < 1$).

Joint entropy

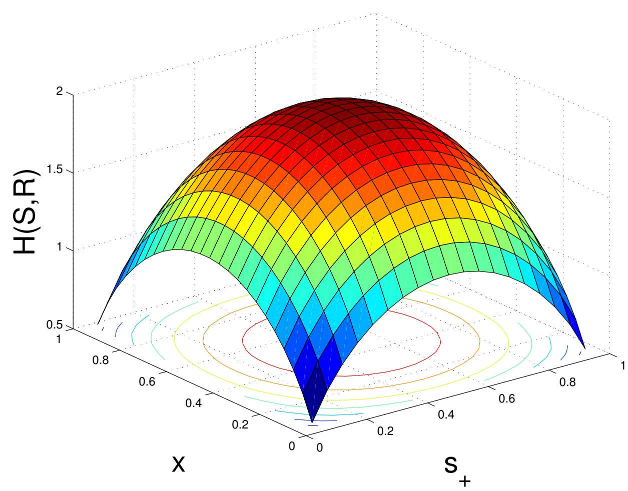

Figure 14: Joint entropy.

Note entropy scale now extends to $2\,bits$.

Mutual information

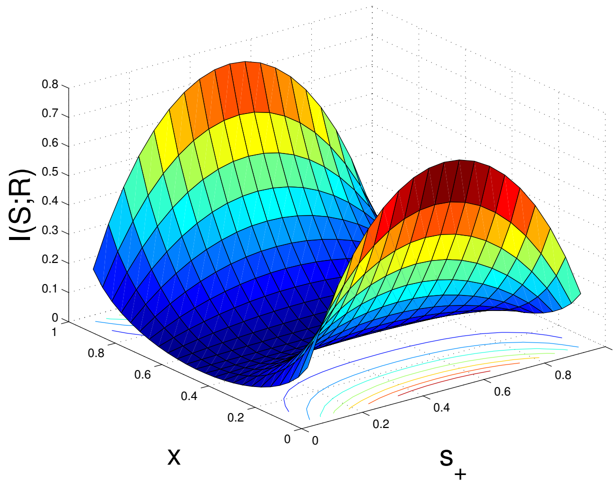

Figure 15: Mutual information.

Mutual information is largest for *deterministic* and smallest for *probabilistic* coupling parameter x.

Summary bivalued stimulus and response

- The example has illustrated some general principles:
- Mutual information is the entropy shared between two dependent events.
- MI is contained within (and thus limited by) both individual entropies.
- MI is maximal for a deterministic and minimal (zero) for a probabilistic dependence.

4 Information of spiking neurons

So far, we have considered only random events with *discrete* outcomes.

Of course, we would like to extend the concept of Shannon entropy to events with *continuous* outcomes.

In doing so, we find that Shannon entropy grows with *measurement precision*, in other words, the number of outcomes that we can reliably distinguish.

We illustrate the issue by computing 'bits per spike' and 'bits per inter-spike-interval' for spiking neurons.

a. Bits per spike

Spike-train entropy calculations are typically based on long recordings with many action potentials. As each spike adds a new 'event', its entropy grows linearly with the recording duration.

Usually, one reports the *rates* with with which entropy or increases with the number of spikes (*bits per spike*) or, alternatively, with time (*bits per second*).

Spikes and samples

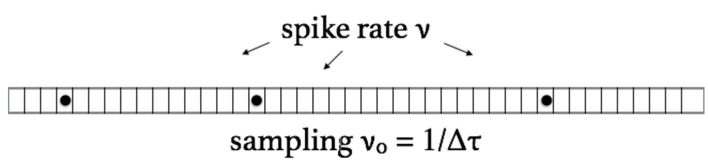

Figure 16: Spikes and samples.

We consider a neuron with firing rate ν, sampled in discrete intervals of rate ν_0. The probability of observing a spike in any particular sample $\Delta t = 1/\nu_0$ is

$$p = \nu\,\Delta t = {}^{\nu}/_{\nu_0}$$

The probability *no* spike is

$$1 - p = 1 - \nu\,\Delta t = {}^{\nu_0 - \nu}\!/_{\nu_0}$$

We ignore the tiny probability of multiple spikes.

Maximal entropy

The entropy is maximal if spike probability is *uniform* over all samples (Poisson spikes).

In this case, the entropy *per sample* Δt, is

$$\begin{aligned} H_{sample} &= p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} = \\ &= \frac{\nu}{\nu_0} \log_2 \frac{\nu_0}{\nu} + \frac{\nu_0 - \nu}{\nu_0} \log_2 \frac{\nu_0}{\nu_0 - \nu} \end{aligned}$$

If spike probability is not uniform, the entropy will be smaller!

Bits per sample

We simplify this result for high sampling rates. In the limit of $\nu_0 \to \infty$, we have

$$\lim_{\nu_0 \to \infty} \left(1 - \frac{\nu}{\nu_0}\right) = 1, \qquad \lim_{\nu_0 \to \infty} \log_2 \left(1 - \frac{\nu}{\nu_0}\right) = -\frac{\nu}{\nu_0} \log_2 e$$

Thus

$$\begin{aligned} H_{sample} &= -\frac{\nu}{\nu_0} \log_2 \frac{\nu}{\nu_0} - \left(1 - \frac{\nu}{\nu_0}\right) \log_2 \left(1 - \frac{\nu}{\nu_0}\right) = \\ &\overset{\nu_0 \to \infty}{=} -\frac{\nu}{\nu_0} \log_2 \frac{\nu}{\nu_0} + \frac{\nu}{\nu_0} \log_2 e = \\ &= \frac{\nu}{\nu_0} \left[- \log_2 \frac{\nu}{e} + \log_2 \nu_0 \right] \end{aligned}$$

Bits per spike

On average, we need to observe

$$N_{sample} = \frac{\nu_0}{\nu}$$

samples $\Delta\tau$ to obtain a spike. Accordingly, the average Shannon information *per spike* is

$$\begin{aligned} H_{spike} &= N_{sample} \, H_{sample} = \\ &= \frac{\nu_0}{\nu} \frac{\nu}{\nu_0} \left[- \log_2 \frac{\nu}{e} + \log_2 \nu_0) \right] \\ &= - \log_2 \frac{\nu}{e} + \log_2 \nu_0 \end{aligned}$$

Here, the first term represents the entropy of firing and the second term the contribution of sampling rate.

Examples

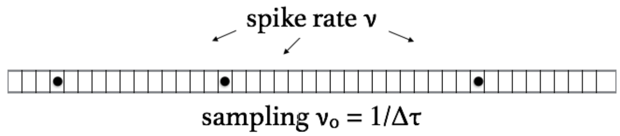

Figure 17: Examples of Spikes and samples.

Sampling rate $\nu_0 = 500\,Hz$, spike rate $\nu = 10\,Hz$

$$H_{spike} = -\log_2 \frac{\nu}{e} + \log_2 \nu_0 = -1.88 + 8.97 = 7.09\,bits$$

Sampling rate $\nu_0 = 2\,kHz$, spike rate $\nu = 30\,Hz$

$$H_{spike} = -\log_2 \frac{\nu}{e} + \log_2 \nu_0 = -3.46 + 10.97 = 7.51\,bits$$

b. Bits per interval

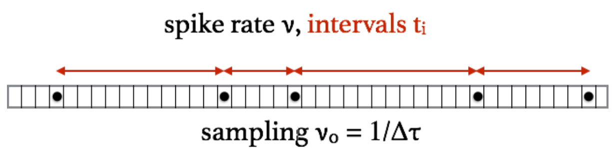

Figure 18: Bits per interval.

To verify the generality of this result, we compare an alternative approach to quantifying the information of spikes.

We define the probability that spikes t_i and t_{i+1} are separated by an interval of $\tau = n\,\Delta\tau$

$$P\left(t_{i+1} - t_i \approx n\,\Delta\tau\right) = p_n\,\Delta\tau$$

and the associated interval density p_n, where $\Delta\tau$ follows again from the sampling rate ν_0

$$\Delta\tau = 1/\nu_0$$

Interval density (optional)

Assuming uniform spike probability (Poisson spikes), and recalling

$$p = \nu\,\Delta\tau = {}^{\nu}\!/_{\nu_0} \approx 0, \qquad\qquad q = 1 - p = {}^{\nu_0 - \nu}\!/_{\nu_0} \approx 1$$

we have

$$\begin{aligned} p_1 &= \frac{p}{\Delta\tau} = \frac{\nu}{\nu_0}\,\nu_0, \qquad\qquad \frac{1}{\nu_0} = \Delta\tau \\ p_2 &= \frac{p\,q}{\Delta\tau} = \nu\,\frac{(\nu_0 - \nu)}{\nu_0} \\ p_3 &= \frac{p\,q^2}{\Delta\tau} = \nu\,\frac{(\nu_0 - \nu)^2}{\nu_0^2}, \\ &\vdots \\ p_n &= \frac{p\,q^{n-1}}{\Delta\tau} = \nu\,\left(1 - \frac{\nu}{\nu_0}\right)^{n-1} \end{aligned}$$

Defining $\tau = n\Delta\tau = \frac{n}{\nu_0}$, we obtain

$$\begin{aligned} p_n &= \nu\,\left(1 - \frac{\nu}{\nu_0}\right)^{n-1} = \\ &= \nu\,\left(1 - \frac{\nu\tau}{n}\right)^{n-1} \\ &\approx \nu\,e^{-n\,\nu/\nu_0} \end{aligned}$$

The approximation is exact in the limit of infinite sampling rate, $\nu_0 \to \infty$

$$p(\tau) = \nu\,e^{-\nu\,\tau}$$

where we have used

$$\lim_{n\to\infty}\left(1 \pm \frac{x}{n}\right)^n = e^{\pm x}$$

Discrete and continuous interval density

$$p_n \approx \nu\, e^{-n\,\nu/\nu_0}, \qquad\qquad p(\tau) = \nu\, e^{-\nu\,\tau}$$

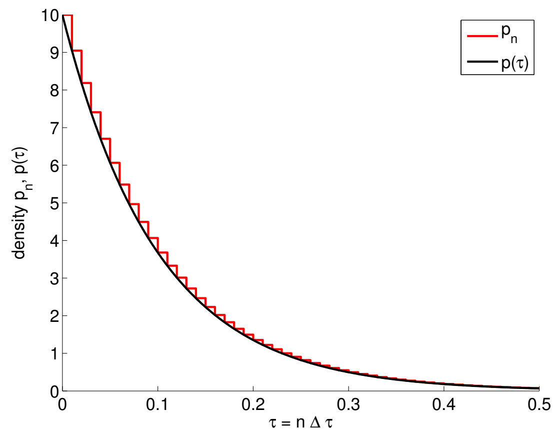

Figure 19: Discrete and continuous interval density

Normalization and mean (optional)

Discrete and continuous density

$$p_n \approx \nu\, e^{-n\,\nu/\nu_0}, \qquad\qquad p(\tau) = \nu\, e^{-\nu\,\tau}$$

Normalization

$$\sum_{k=0}^{\infty} p_k\, \Delta\tau \approx 1, \qquad\qquad \int_0^{\infty} p(\tau)\, d\tau = 1$$

Expectation

$$\sum_{k=0}^{\infty} \frac{k}{\nu_0}\, p_k\, \Delta\tau \approx \frac{1}{\nu}, \qquad\qquad \int_0^{\infty} \tau\, p(\tau)\, d\tau = \frac{1}{\nu}$$

Entropy per interval (optional)

By definition, the Shannon entropy per interval is

$$\begin{aligned} H_{isi} &= \sum_n p_n \, \Delta\tau \, \log_2 \frac{1}{p_n \, \Delta\tau} = \\ &= \Delta\tau \sum_n p_n \, \log_2 \frac{1}{p_n} + \Delta\tau \, \log_2 \frac{1}{\Delta\tau} \sum_n p_n = \\ &= \Delta\tau \sum_n p_n \, \log_2 \frac{1}{p_n} + \log_2 \nu_0 \end{aligned}$$

since

$$\sum_n p_n \, \Delta\tau = 1, \qquad \frac{1}{\Delta\tau} = \nu_0$$

In the continuous limit,

$$\begin{aligned} H_{isi} &= \int_0^\infty p(\tau) \, \log_2 \frac{1}{p(\tau) \, d\tau} \, d\tau = \\ &= \int_0^\infty p(\tau) \, \log_2 \frac{1}{p(\tau)} \, d\tau - \log_2 \, d\tau \int_0^\infty p(\tau) \, d\tau = \\ &= \int_0^\infty p(\tau) \, \log_2 \frac{1}{p(\tau)} \, d\tau - \log_2 \, d\tau \end{aligned}$$

since

$$\int_0^\infty p(\tau) \, d\tau = 1, \qquad \int_0^\infty \tau \, p(\tau) \, d\tau = \frac{1}{\nu}$$

Poisson spikes

In case of uniform spike probability (see above) with

$$p(\tau) = \nu\, e^{-\nu\,\tau} = \nu\, 2^{-\nu\,\tau\,\log_2 e}$$

the entropy per spike is

$$\begin{aligned} H_{spike} &= \int_0^\infty p(\tau)\, \log_2 \frac{1}{p(\tau)}\, d\tau - \log_2\, \Delta\tau = \\ &= \int_0^\infty p(\tau)\, [-\log_2 \nu + \tau\,\nu\, \log_2 e]\, d\tau - \log_2\, \Delta\tau = \\ &= -\log_2 \nu + \frac{\nu}{\nu}\, \log_2 e - \log_2\, \Delta\tau = \\ &= -\log_2 \frac{\nu}{e} + \log_2 \nu_0 \end{aligned}$$

This is identical to the result obtained by considering sampling intervals Δt! Shannon information describes the **process** itself, **not the description** we choose to adopt.

c. Multiple spikes

A single spike with average rate ν has Shannon entropy

$$H_1 = -\log_2 \frac{\nu}{e} + \log_2 \nu_0$$

This assumes uniform probability (Poisson spikes). Any deviation *decreases* entropy/information!

N spikes with average rate ν have Shannon entropy

$$H_N = -N\, \log_2 \frac{\nu}{e} + N\, \log_2 \nu_0$$

This assumes independence. Any deviation *decreases* entropy/information!

Less Shannon information

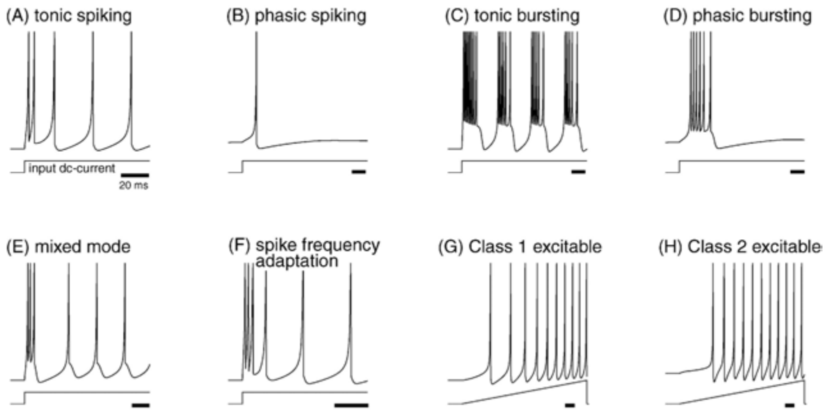

Figure 20: Less Shannon information. [1]

More Shannon information

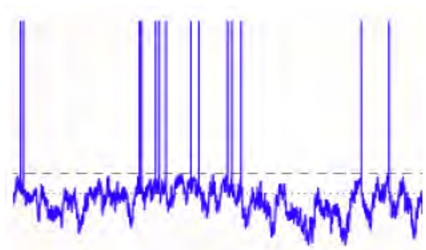

Figure 21: More Shannon information.

Summary information of spiking neurons

- The Shannon entropy of individual is limited by spike rate ν and sampling rate ν_0. For uniform spike probability (Poisson spikes) it is

$$H_{spike} = -\log_2 \frac{\nu}{e} + \log_2 \nu_0$$

- The Shannon entropy of multiple spikes is subject to the same limit. For Poisson spikes

$$H_{N\,spikes} = -N\ \log_2 \frac{\nu}{e} + N\ \log_2 \nu_0$$

- This holds for spikes from either single or multiple neurones. It is independent of how spike probability is computed (single, pair, triplet, ...).

- Any deviation from uniformity and independence *decreases* Shannon information/entropy (firing patterns, synfire chains, synchronisation, ...)

5 Neurons with variable firing rates

We consider the entropy of neurons with variable firing rates. As always, the entropy will depend on how many rates there are and on how frequently they are observed.

Accordingly, the entropy depends on the precision with which rates can be established. In addition, it depends on how large the firing rates are allowed to be.

Thus, the entropy depends crucially on constraints such as the maximum, the average, or the variability of firing rates.

We will use information theory to reveal the implications of such constraints!

Maximum constrained

If the maximal response (firing rate) of a neuron is fixed, we seek the response distribution $p(r)$ that maximizes response entropy

$$H(R) = \int_0^{r_{max}} p(r)\, \log_2 \frac{1}{p(r)}\, dr$$

subject to

$$\int_0^{r_{max}} p(r)\, dr = 1$$

The solution is a *flat* distribution

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

$$H = \log_2 r_{max} - \log_2 \Delta r = \log_2 \frac{r_{max}}{\Delta r}$$

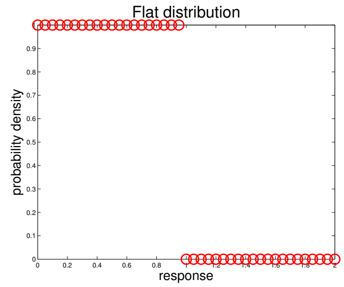

Figure 22: Flat distribution.

Mean constrained

If we fix the mean response instead, we seek the $p(r)$ that maximizes

$$H(r) = \int_0^\infty p(r)\, \log_2 \frac{1}{p(r)}\, dr$$

subject to

$$\langle r \rangle = \int_0^\infty r\, p(r)\, dr = r_0, \qquad \int_0^\infty p(r)\, dr = 1$$

The solution is an *exponential* distribution

$$p(r) = \frac{1}{r_0} \exp\left(-\frac{r}{r_0}\right)$$

$$H = \log_2 \frac{e\, r_0}{\Delta r}$$

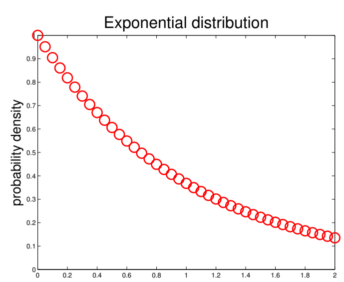

Figure 23: Exponential distribution.

Variance constrained

Finally, if we prescribe both mean and variance of the response, we seek the $p(r)$ that maximizes

$$H(R) = \int_{-\infty}^{+\infty} p(r) \, \log_2 \frac{1}{p(r)} \, dr$$

subject to

$$\langle (r - r_0)^2 \rangle = \sigma_0^2, \qquad \langle r \rangle = r_0, \qquad \int_{-\infty}^{+\infty} p(r) \, dr = 1$$

The solution is a *Gaussian* distribution

$$p(r) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp\left(-\frac{(r - r_0)^2}{2\sigma_0^2} \right)$$

$$H = \log_2 \frac{\sigma_0}{\Delta r} + \frac{1}{2} \log_2 (2\pi e)$$

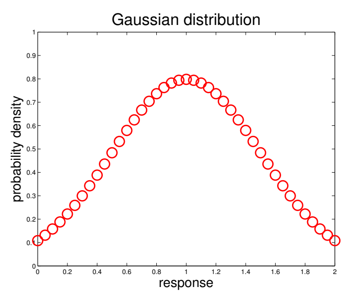

Figure 24: Gaussian distribution.

Summary neurons with variable firing rates

- The Shannon entropy depends on how many firing rates we can distinguish and on how often they are observed.

- Entropy is maximised by different distributions $P(r)$, depending on constraints:
  - Maximum constrained $\rightarrow$ flat
  - Average constrained $\rightarrow$ exponential
  - Variance constrained $\rightarrow$ Gaussian

- Information theory predicts that response distributions reflect design constraints of firing mechanisms and/or neuronal metabolism.

6. Histogram equalization

Suppose entropy is maximized by a certain response distribution $p(r)$ (which depends on constraints).

How can a neuron ensure that its actual response distribution approximates the optimal distribution?

After all, responses must reflect the stimulus distribution $p(s)$, which reflects the environment!

The technique used by neurons to ensure that a given stimulus distribution $p(s)$ results in a given (optimal) response distribution $p(r)$ is **histogram equalization**!

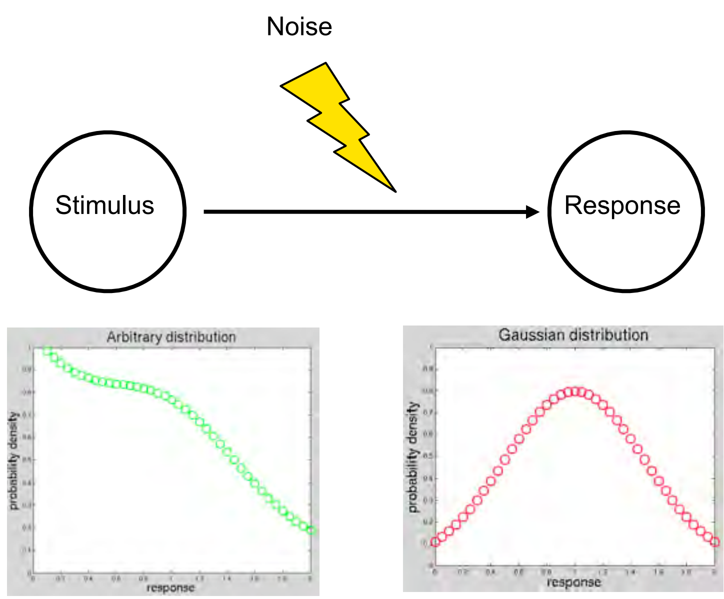

Figure 25: Histogram equalization.

The method of *histogram equalization* involves choosing the right *transfer function* $r = f(s)$!

A suitable transfer function

$$r = f(s) \quad \text{monotonically increasing}$$

must satisfy

$$p_s(s)\, ds = p_r\, [f(s)]\, dr \qquad \Leftrightarrow \qquad \int_{s_{min}}^{s} p_s(s)\, ds = \int_{r_{min}}^{r=f(s)} p_r(r)\, dr$$

A transfer function $f(s)$ satisfying this condition ensures that a given stimulus distribution $p(s)$ produces the desired response distribution $p(r)$.

Key idea of histogram equalization

Local slope of transfer function

$$\frac{dr}{ds} = f'(s)$$

determines, for a given value range Δs, to how large a value range of Δr it is mapped. In this way, the transfer function can 'equalize' the cumulative probabilities

$$P(s \in [s, s + \Delta s]) = P(r \in [r, r + \Delta r])$$

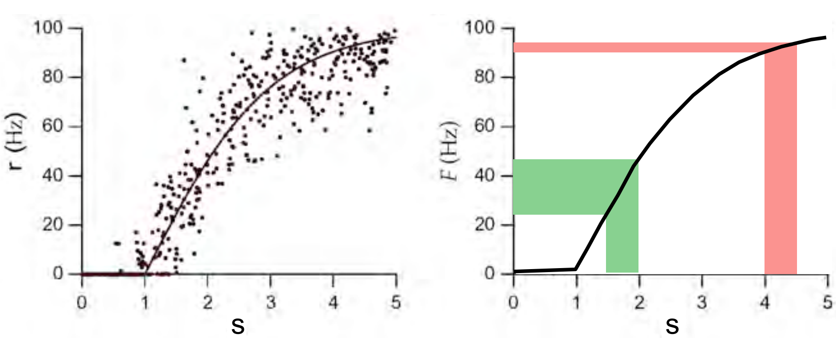

Figure 26: Transfer function.

Flat p(r), arbitrary p(s), seek r=f(s)

Consider a neuron with a fixed *maximal* response r_{max}, so that entropy is maximized by a flat response probability

$$p_r(r) = \frac{1}{r_{max} - r_{min}}$$

$$\begin{aligned}\int_{s_{min}}^{s} p_s(s)\, ds &= \int_{r_{min}}^{r=f(s)} p_r(r)\, dr = \\ &= \frac{1}{r_{max} - r_{min}} \int_{r_{min}}^{r=f(s)} dr = \frac{f(s) - r_{min}}{r_{max} - r_{min}}\end{aligned}$$

$$f(s) = r_{min} + (r_{max} - r_{min}) \int_{s_{min}}^{s} p_s(s)\, ds$$

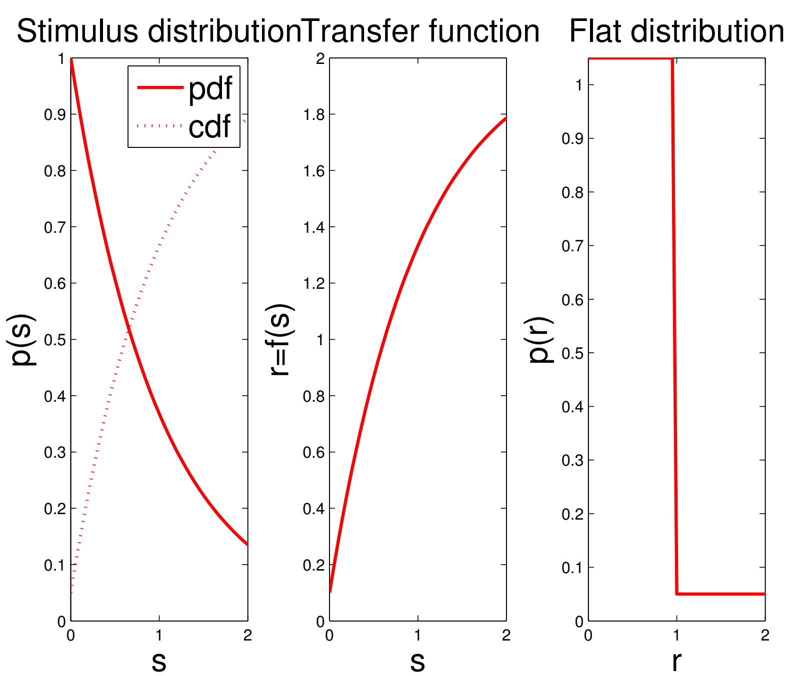

Figure 27: Neuron with fixed *maximal* response 1.

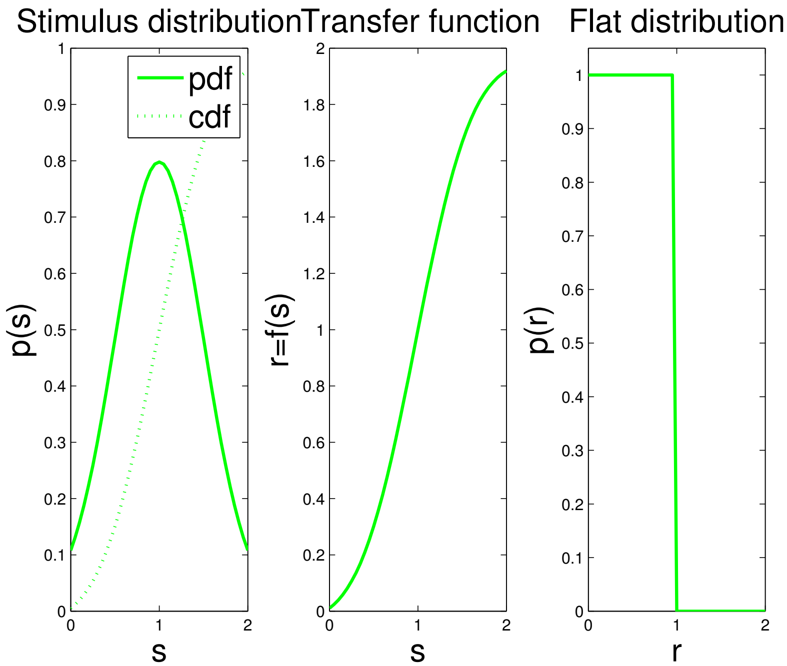

Figure 28: Neuron with fixed *maximal* response 2.

Exponential p(r), arbitrary p(s), seek r=f(s)

Consider a neuron with a fixed *average* response r_0, so that entropy is maximized by an exponential response probability

$$p_r(r) = \frac{1}{r_0} \exp\left(-\frac{r}{r_0}\right)$$

$$\begin{aligned}\int_{s_{min}}^{s} p_s(s)\, ds &= \int_{r_{min}}^{r=f(s)} p_r(r)\, dr = \\ &= \left[-\exp\left(-\frac{r}{r_0}\right)\right]_0^{r=f(s)} = 1 - \exp\left(-\frac{f(s)}{r_0}\right)\end{aligned}$$

$$f(s) = -r_0 \ln\left[1 - \int_{s_{min}}^{s} p_s(s)\, ds\right]$$

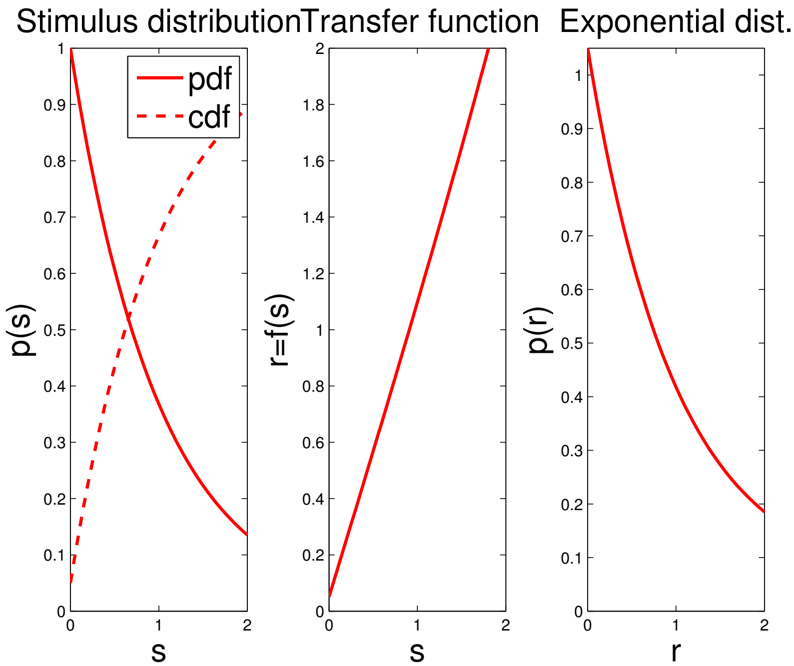

Figure 29: Neuron with a fixed *average* response 1.

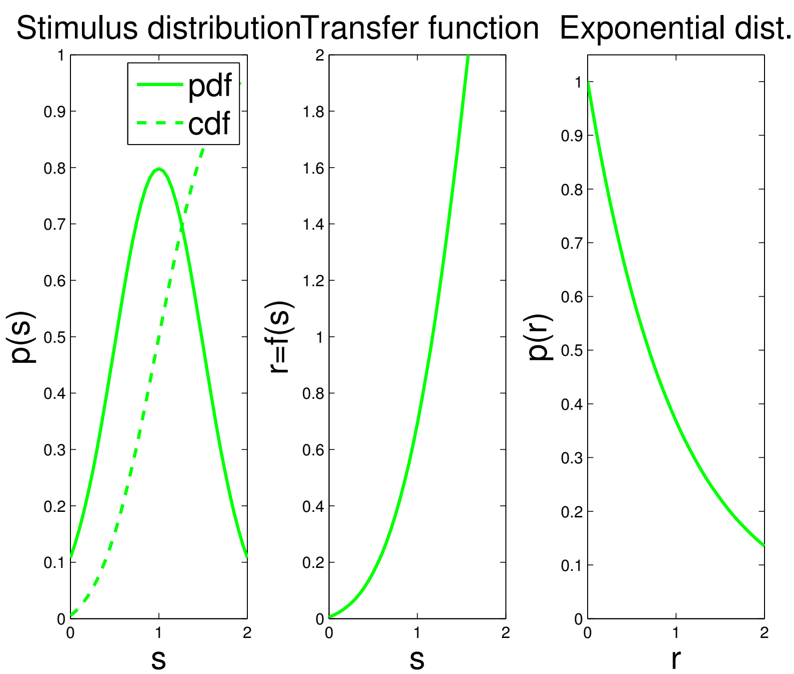

Figure 30: Neuron with a fixed *average* response 2.

Gaussian p(r), arbitrary p(s), seek r=f(s)

Consider a neuron constrained to a Gaussian response probability

$$p_r(r) = \frac{1}{\sqrt{2\pi\sigma_r}} \exp\left(-\frac{(r - r_0)^2}{2\sigma_r^2}\right)$$

$$\int_{-\infty}^{r} p_r(r)\, dr = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{r - r_0}{\sqrt{2}\sigma_r}\right)$$

$$\int_{-\infty}^{s} p_s(s)\, ds = \int_{-\infty}^{r=f(s)} p_r(r)\, dr = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{f(s) - r_0}{\sqrt{2}\sigma_r}\right)$$

$$f(s) = r_0 + \sqrt{2}\,\sigma_r\,\mathrm{erf}^{-1}\left[2\int_{-\infty}^{s} p_s(s)\, ds - 1\right]$$

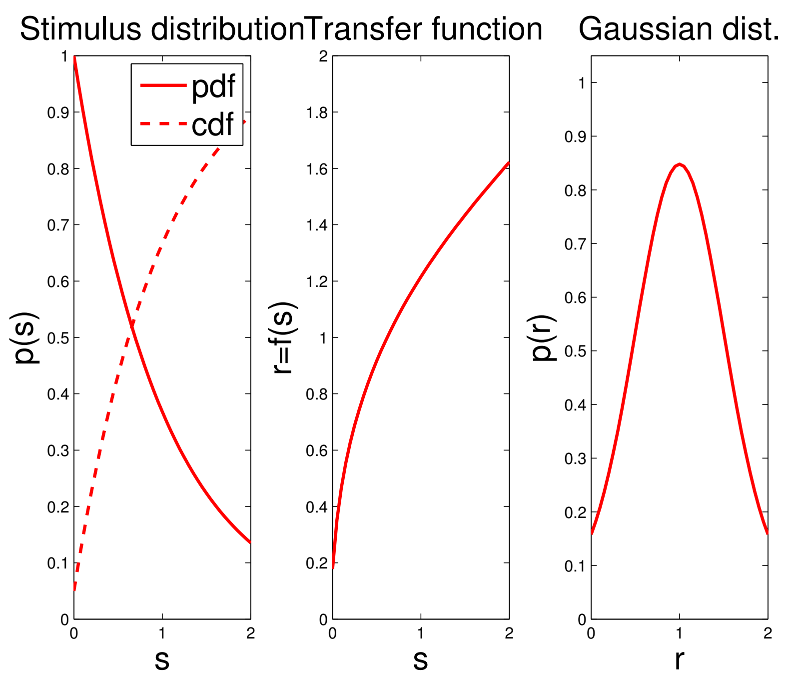

Figure 31: Neuron constrained to a Gaussian response probability.

Special case: if the stimulus distribution is Gaussian, too, the optimal transducer is linear. In this case, we can match stimulus and response distributions by appropriate shifting and scaling.

$$p_s(s) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left(-\frac{(s-s_0)^2}{2\sigma_s^2}\right)$$

$$\int_{-\infty}^{s} p_s(s)\, ds = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{s-s_0}{\sqrt{2}\sigma_s}\right)$$

$$f(s) = r_0 + \frac{\sigma_r}{\sigma_s}(s - s_0)$$

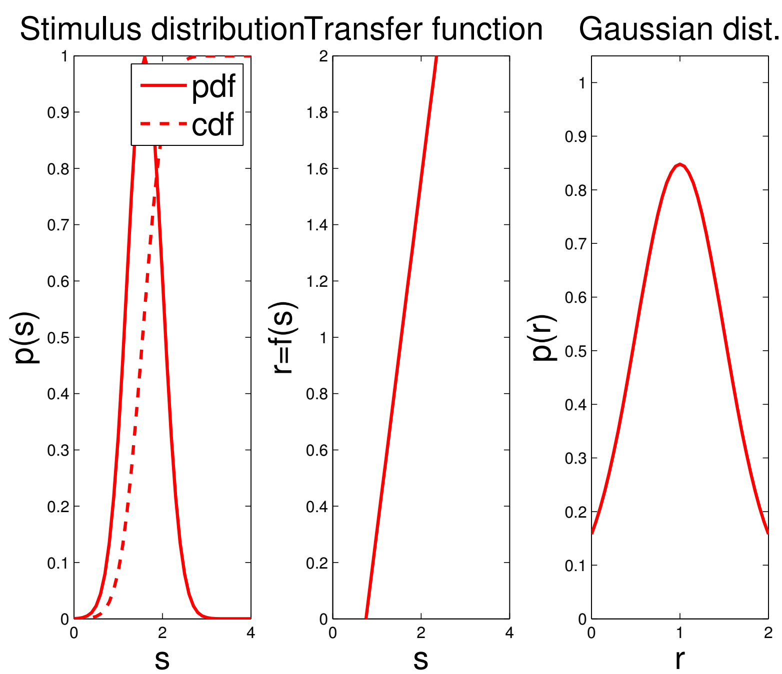

Figure 32: Special case 1.

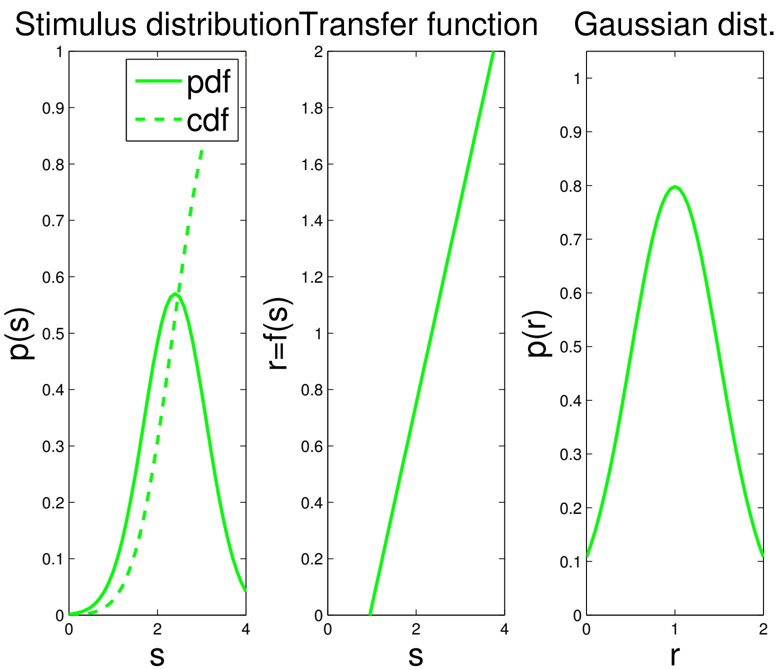

Figure 33: Special case 2.

Large monopolar cell of the fly

Neurons do use histogram equalization! The first reported case was the 'large monopolar cell' (LMC) in the fly (Laughlin, 1981), which responds differentially to different levels of image contrast. Laughlin measured the probability of contrast levels in natural scenes from the habitat of the fly and computed the integral over contrast. The result almost exactly coincided with the measured transducer function $r = f(s)$.

$$f(c) = r_{min} + (r_{max} - r_{min}) \int_{c_{min}}^{c} p_c(c)\, dc$$

Thus, the LMC seems to maximize response entropy under the design constraint of a 'flat' response distribution.

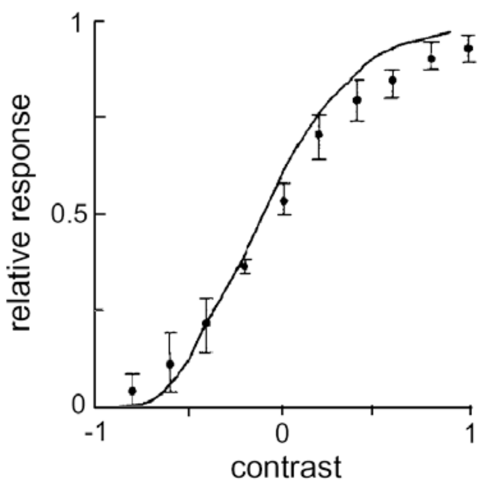

Figure 34: Large monopolar cell of the fly. [2]

Line: intergral over $p(c)$. mean±std.err.: measured $f(c)$

Summary histogram equalization

- Neuronal responses reflect the frequency of different stimuli in the natural environment.
- To effectively use the response range available, a suitable transfer function maps frequent stimuli to a larger response range than infrequent stimuli.
- In this way, arbitrary stimulus distributions are transformed into maximal-entropy response distributions.
- Thus, sensory systems must incorporate information about the natural environment!
- Information theory links seemingly unrelated facts (*e.g.*, stimulus statistics and transfer functions).

Overall summary

- Flat distributions are maximally informative!
- Optimal systems communicate with flat distributions!
- Poisson spikes are equally likely at any time (flat distribution)!
- Neural communication consistent with information theory.

6 Bibliography

1. Dayan & Abbott (2001) Theoretical neuroscience, MIT Press.
2. Dayan & Abbott (2001) Theoretical neuroscience, MIT Press. Fig 4.2