Lecture 9

2025-10-29

Probability; Encoding and decoding; Bayesian inference; Test size and power; Receiver operating characteristic (ROC) curve; Introduction to information theory

- Online slides: lec09-probability_inference.html
- Code: code09.ipynb
- Homework: hw09.pdf, hw09-data.npz

Username: cns, Password: nycu2025

Laboratory of Computing Biological Networks

Modeling with Stochasticity

Noise is present in the models to, e.g.,

- account for variability in response
- capture statistical properties of complex environment

Some probability related concepts

- Conditional probability
- Joint probability
- Prior probability
- Marginal probability
- Bayes' theorem

Distribution of two random variables

Consider random variables x and y:

Joint probability $P(x, y)$

The chance of a given pair x, y of observations.

Normalization: $\sum_{x,y} P(x, y) = 1$

Marginal probabilities $P(x)$, $P(y)$

$P(x) = \sum_y P(x, y)$ (Don't care about ⇒ Sum over it.)

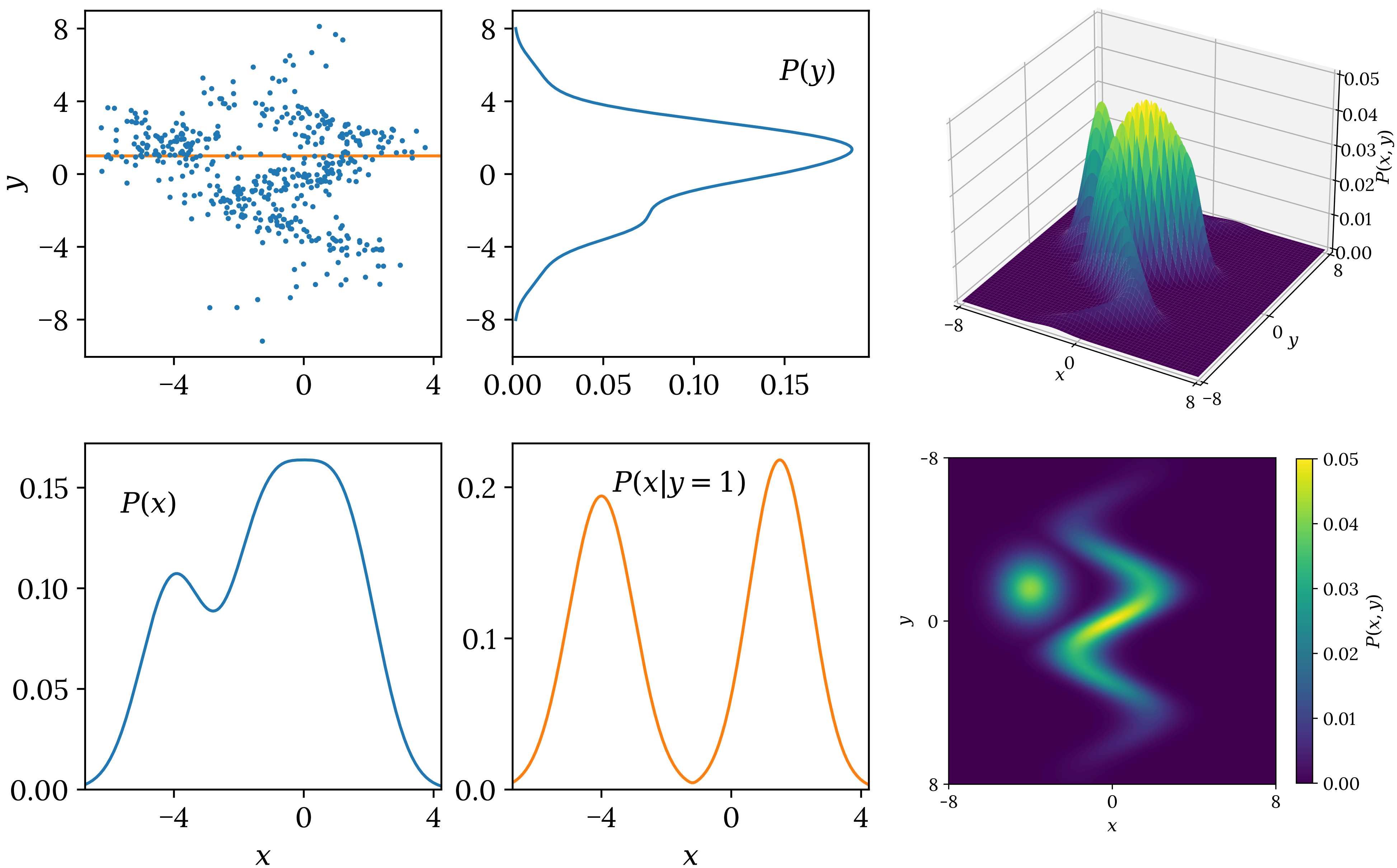
Conditional probability $P(x|y)$

$P(x|y)P(y) = P(x, y)$, distribution of x knowing y

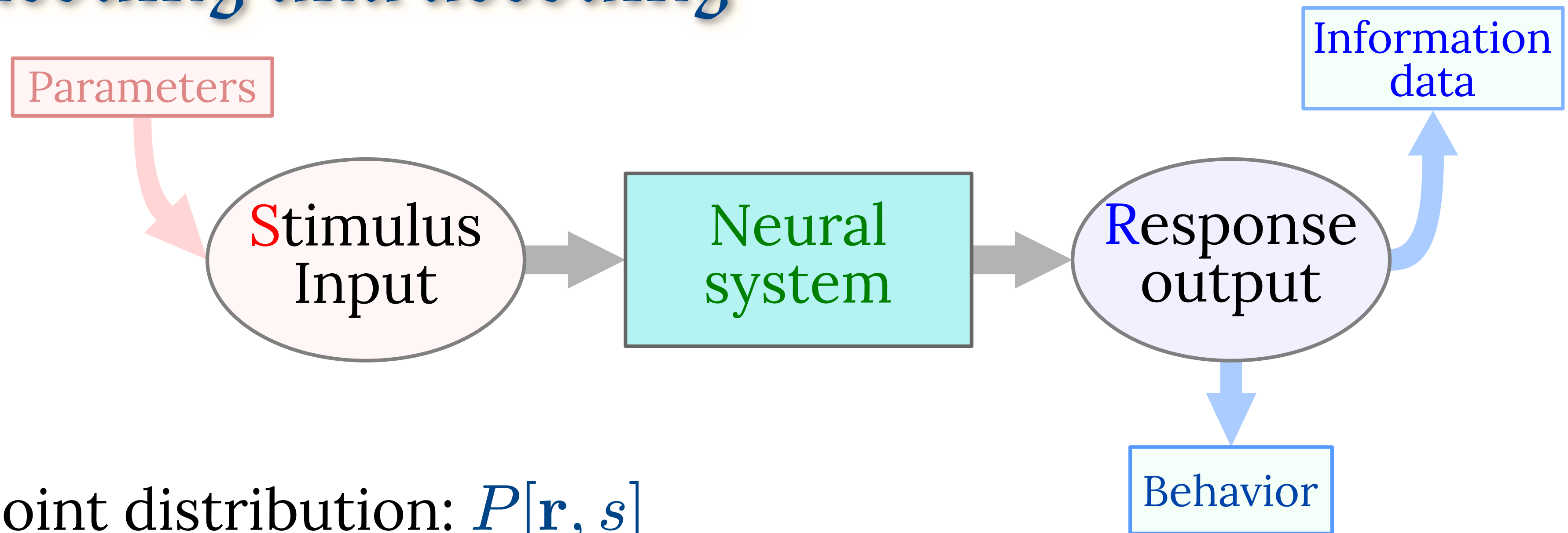
Bayesian prior & posterior

Expected distribution of x before and after measuring y ⇒ $P(x)$ and $P(x|y)$

Illustration of two variable distribution



Encoding and decoding



- <u>Joint distribution</u>: $P[\mathbf{r}, s]$
  The whole truth (consider s and $\mathbf{r}$ together as *the* variable)
- <u>Marginal distribution</u>: $P[\mathbf{r}]$, $P[s]$
  Only seeing half of the world
- <u>Conditional distributions</u>: $P[s|\mathbf{r}]$, $P[\mathbf{r}|s]$
  Fixing one and seeing the other

$P[\mathbf{r}] = \sum_s P[\mathbf{r}|s]P[s]$ and similarly $P[s] = \sum_{\mathbf{r}} P[s|\mathbf{r}]P[\mathbf{r}]$

Bayesian inference

$$P[\mathbf{r}, s] = P[\mathbf{r}|s]P[s] = P[s|\mathbf{r}]P[\mathbf{r}]$$

Bayes theorem

$$P[s|\mathbf{r}] = \frac{P[\mathbf{r}|s]P[s]}{P[\mathbf{r}]}$$

Inferring the stimulus s from an observed response $\mathbf{r}$.

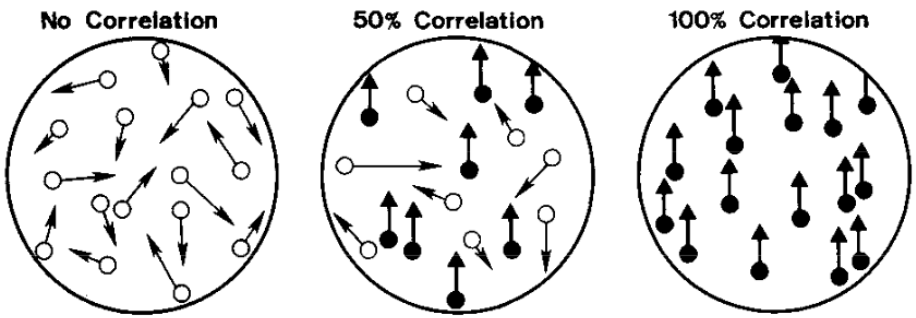
$\downarrow s,\ \mathbf{r} \rightarrow$	**Positive**	**Negative**
Infected	96%	4%
Safe	2%	98%

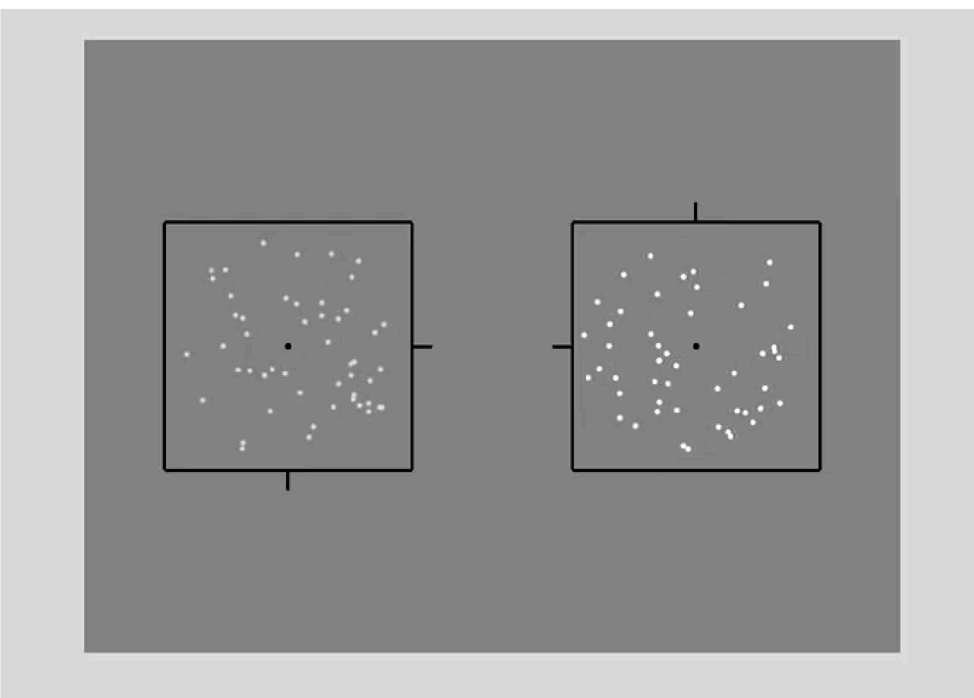
If tested positive, how much should I worry about it?

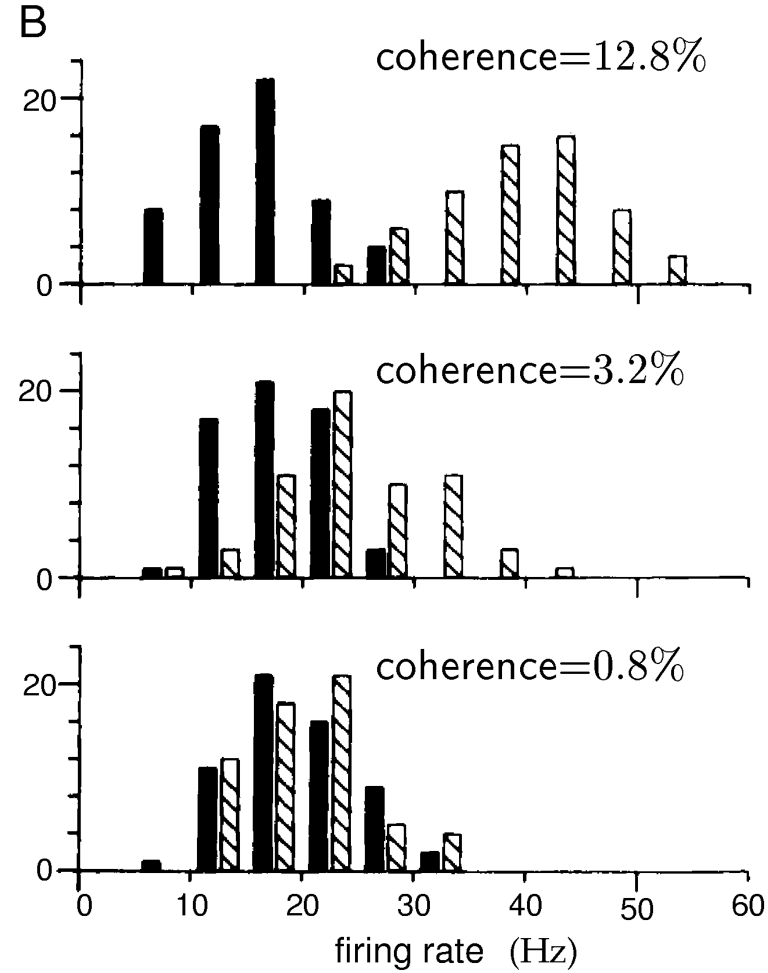
The lazy pollster

Consider a referendum with a choice of two options **A** *and* **B**. A *pollster was tasked with estimating the likely outcome by conducting a poll. Suppose the pollster was so lazy that they simply got on the street and asked just the first person they saw. If they received* **B** *as the response, what is the estimated percentage of the outcome* **A** *that the pollster should report?*

Moving random dot stimuli







https://www.youtube.com/watch?v=11564eLhSpo

A rat version: https://www.youtube.com/watch?v=oDxcyTn-0os.

Discriminability:

$$d' = \frac{\langle \mathbf{r} \rangle_+ - \langle \mathbf{r} \rangle_-}{\sigma_{\mathbf{r}}}$$

Dayan & Abbott (3.4)

Test size and power

$\alpha(z) = P[r \geq z|-]$ size or false alarm rate

$\beta(z) = P[r \geq z|+]$ power or hit rate

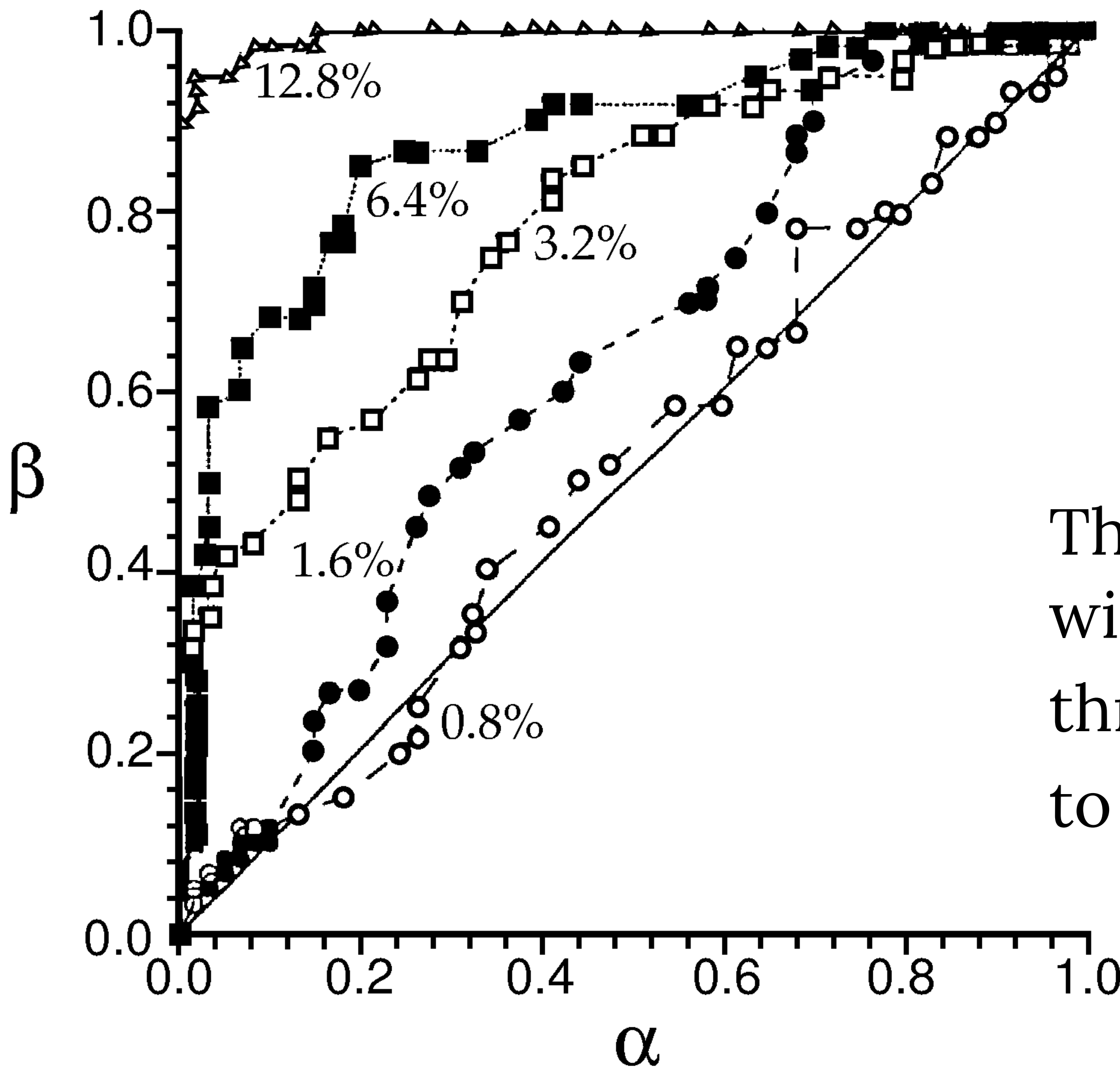
"z" is the decision threshold

	Probability	
Stimulus	**Correct**	**Incorrect**
$+$	β	$1-\beta$
$-$	$1-\alpha$	α

Would like to maximize the probability of correct answer:

$$\beta P[+] + (1-\alpha)P[-]$$

Receiver operating characteristic (ROC) curve



The curves are traced out with the value of decision threshold z goes from $-\infty$ to $+\infty$.

Two-alternative forced choice

Two tests of opposite stimuli force a choice without given z.

Consider the stimulus $s_1 = +$ and use r_2 as the threshold z

$$P[\text{correct}] = \int_0^\infty dz\, p[z|-]\beta(z)$$

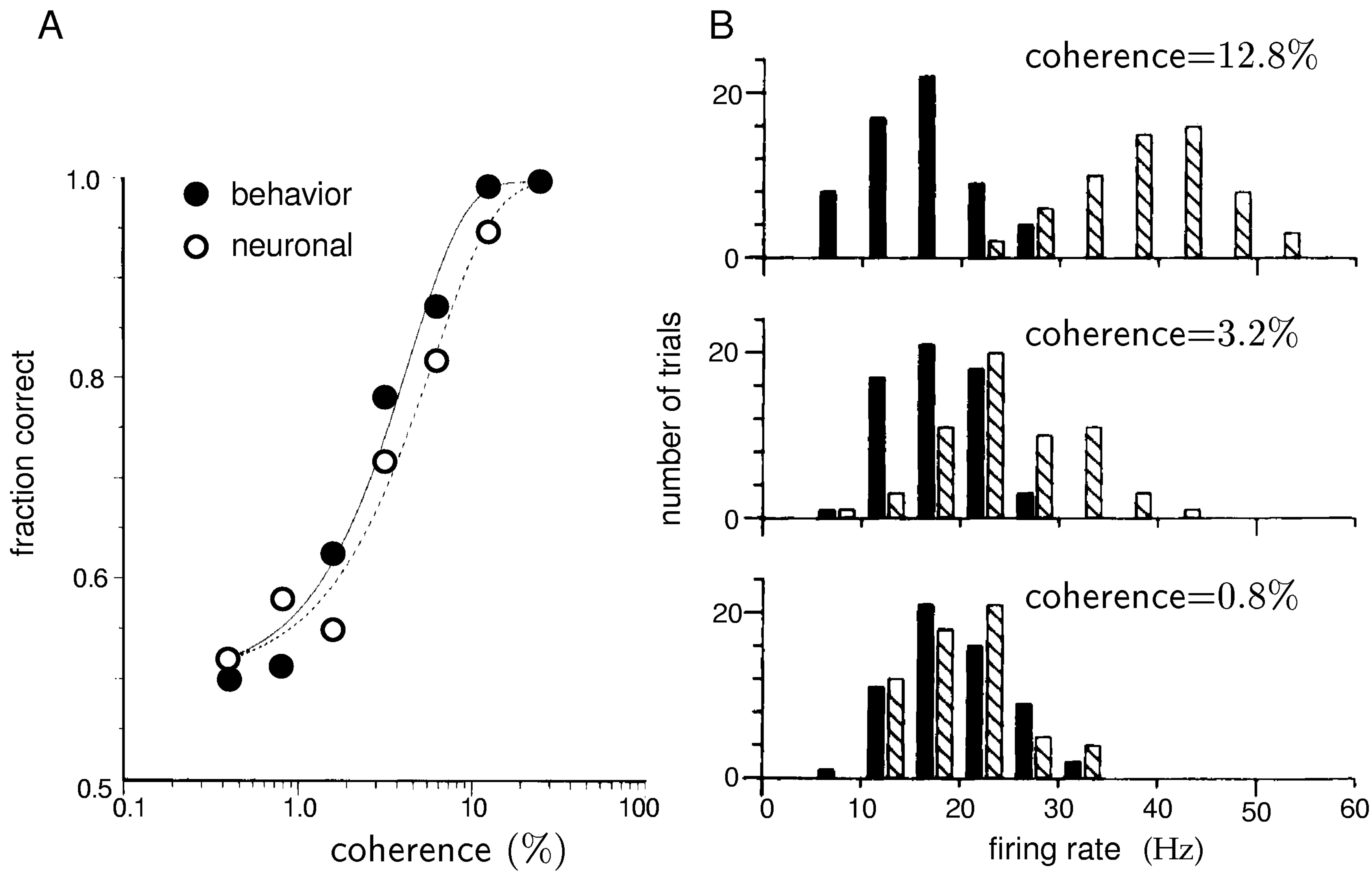
$$\alpha(z) = \int_z^\infty dr\, p[r|-] \quad \Rightarrow \quad \frac{d\alpha}{dz} = -p[z|-]$$

We arrive at

$$\Rightarrow P[\text{correct}] = \int_0^1 d\alpha \beta$$

which is the area under the ROC curve.

Fractions of correct inference



Concepts in information theory

Points to note:

- What is information?
- How to measure information?
- Conditional probability
- Bayesian and Prior
- Expectation and most likelihood

What is entropy

- Entropy measures how "*interesting*" or "*surprising*" a set of responses is.
  — Dayan & Abbott

- Entropy is the number of possible configurations of a system's components that is consistent with the state of the system as a whole.
  — Wikipedia

- Entropy is the degree of disorder or uncertainty in a system
  — Merriam-Webster

Shannon information (entropy)

Consider information h obtained from an observation result $\mathbf{r}$ with probability $P[\mathbf{r}]$:

- Information of r should decrease with an increase of $P[\mathbf{r}]$.
- For independent observations $\mathbf{r}_1$ and $\mathbf{r}_2$:

$$P[\mathbf{r}_1, \mathbf{r}_2] = P[\mathbf{r}_1]P[\mathbf{r}_2]$$

- The total information is a sum of the two:

$$h(P[\mathbf{r}_1]P[\mathbf{r}_2]) = h(P[\mathbf{r}_1]) + h(P[\mathbf{r}_2])$$

- Only choice (up to a constant factor):

$$h(P[\mathbf{r}]) = -\log_2 P[\mathbf{r}]$$

Shannon entropy of a distribution

$$H = -\sum_{\mathbf{r}} P[\mathbf{r}] \log_2 P[\mathbf{r}]$$

- For discrete variable $\mathbf{r}$.
- For a rare event $P[\mathbf{r}] = \epsilon \to 0$, $h(P[\mathbf{r}])$ diverges but we still have $\epsilon \log \epsilon \to 0$
- Taking continuous limit of $\mathbf{r}$ leads to divergence in H.
- For finite number of states, H maximizes when $P[\mathbf{r}]$ is constant for all $\mathbf{r}$.

Noise in the response

- Information (from a response) is useful only when it tells us about the things we care (the stimulus).

Entropy of response to a given stimulus:

$$H_s = -\sum_{\mathbf{r}} P[\mathbf{r}|s] \log_2 P[\mathbf{r}|s]$$

Average over stimuli gives entropy of response noise:

$$H_{\text{noise}} = \sum_{s} P[s] H_s = -\sum_{s,\mathbf{r}} P[s] P[\mathbf{r}|s] \log_2 P[\mathbf{r}|s]$$

Mutual information

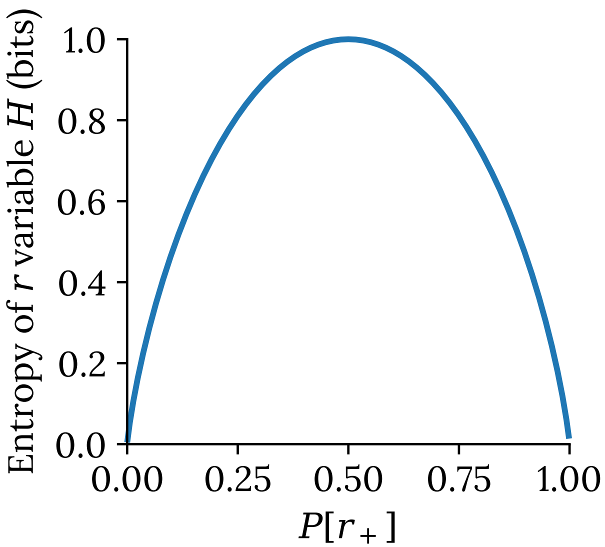
Information about the stimulus:

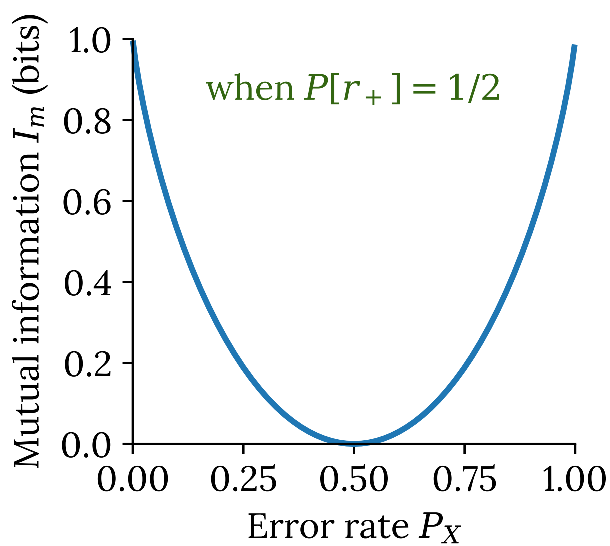
$$
\begin{aligned}
I_m &= H - H_{\text{noise}} \\
&= -\sum_{\mathbf{r}} P[\mathbf{r}] \log_2 P[\mathbf{r}] + \sum_{s,\mathbf{r}} P[s] P[\mathbf{r}|s] \log_2 P[\mathbf{r}|s] \\
&= \sum_{s,\mathbf{r}} P[s] P[\mathbf{r}|s] \log_2 \frac{P[\mathbf{r}|s]}{P[\mathbf{r}]} \\
&= \sum_{s,\mathbf{r}} P[\mathbf{r}, s] \log_2 \frac{P[\mathbf{r}, s]}{P[\mathbf{r}] P[s]}
\end{aligned}
$$

Symmetric between s and $\mathbf{r}$!

Example: two-state stimulus and response

- Entropy of stimulus and response are both 1 bit when probability of two states are equal.
- MI goes to 1 for both $P_X = 0$ and 1 and becomes 0 when $P_X = 1/2$.





	r_+	r_-
s_+	$(1 - P_X)/2$	$P_X/2$
s_-	$P_X/2$	$(1 - P_X)/2$

Mutual information is:

$$I_m = 1 + (1 - P_X)\log_2(1 - P_X) + P_X \log_2 P_X$$