CSE 528

1/27/2017

Today:

+ signal detect. thry
+ decoding from mult. neurons
+ inf. thry

Review of Bayes' Law: $p(\vec{r},\vec{s}) = p(\vec{r}|\vec{s})p(\vec{s}) = p(\vec{s}|\vec{r})p(\vec{r})$

joint, conditional, marginal

$$\underbrace{p(\vec{s}|\vec{r})}_{\text{posterior}} = \frac{\underbrace{p(\vec{r}|\vec{s})}_{\text{likelihood}}\underbrace{p(\vec{s})}_{\text{prior}}}{p(\vec{r})}$$

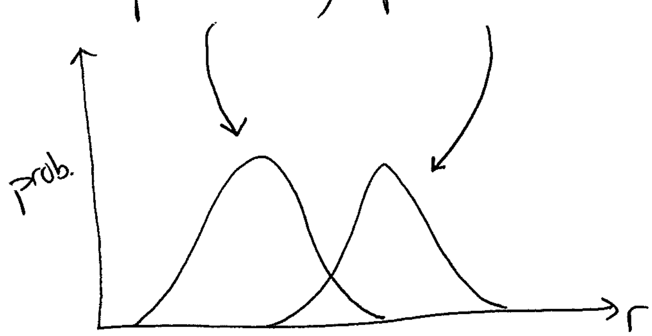
Signal Detection Theory

Goal: given a ~~binary stimulus~~ neural response to a binary stim, determine if stim is 0 or 1

E.g., $s = 0$ no tiger
$\quad\;\; = 1$ tiger

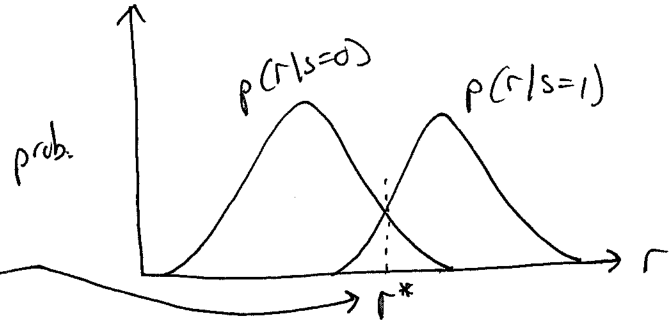
$r \in [0, r_{max}]$ firing rate

Relevant quantities: prior: $p(s=0)$, $p(s=1) = 1 - p(s=0)$
likelihood: $p(r|s=0)$, $p(r|s=1)$



Specific goal: figure out rule for determining if $s=0$ or 1 given r, i.e., choose threshold r^* s.t. if $r > r^*$ guess 1
if $r < r^*$ guess 0

Simplest s|n:



$\boxed{\text{max. likelihood}}$ (ML): choose 1 when $p(r|s=1) > p(r|s=0)$

0 otherwise $\longrightarrow$ r^* is s|n to $p(r|s=1) = p(r|s=0)$

false pos. rate = $p(s=0)\,p(\text{choose } 1|s=0) = p(s=0)\,p(r > r^*|s=0)$

$$= p(s=0)\int_{r^*}^{\infty} p(r|s=0)\,dr$$

false neg. rate = $p(s=1)\,p(\text{choose } 0|s=1) = p(s=1)\,p(r < r^*|s=1)$

$$= p(s=1)\int_{-\infty}^{r^*} p(r|s=1)\,dr$$

problem: consider when $r = r^*$ $\longrightarrow$ most of the time $s=0$ (no tiger)

$\boxed{\text{ML neglects prior}}$

better s|n: choose 1 when $p(s=1|r) > p(s=0|r)$

$\boxed{\text{Maximum a posteriori}}$ (MAP) i.e., $\dfrac{p(s=1)p(r|s=1)}{\cancel{p(r)}} > \dfrac{p(s=0)p(r|s=0)}{\cancel{p(r)}}$

this incorporates prior probability $\rightarrow$ if $p(s=1)$ very small, then $p(r|s=1)$ must be very large (rel. to $p(r|s=0)$) to choose 1

$\hookrightarrow$ r^* is s|n to:

$$p(s=1)p(r|s=1) = p(s=0)p(r|s=0)$$

i.e., $\dfrac{p(r|s=1)}{p(r|s=0)} = \dfrac{p(s=0)}{p(s=1)}$

alternative s|n: consider cost of guessing incorrectly

$C(\text{choose } 1, s=0) \leftarrow$ guess tiger but no tiger (low cost)

$C(\text{choose } 0, s=1) \leftarrow$ guess no tiger but tiger! (high cost)

in this case:

given r, choose 1 when $E_s[C(\text{choose } 1, s)|r] < E_s[C(\text{choose } 0, s)|r]$

0 otherwise

$$E_s[C(\text{choose } 1, s)|r] = p(s=0|r)\,C(\text{choose } 1|s=0) + p(s=1|r)\,\underbrace{C(\text{choose } 1|s=1)}_{0}$$

0 (assuming no cost for correct guess)

$$= p(s=0|r)\,C(\text{choose } 1|s=0)$$

$$= \frac{p(s=0)\,p(r|s=0)}{p(r)}\,C(\text{choose } 1|s=0)$$

$$E_s[C(\text{choose } 0, s)|r] = \frac{p(s=1)\,p(r|s=1)}{p(r)}\,C(\text{choose } 0|s=1)$$

i.e., choose 1 when $\frac{p(s=0)\,p(r|s=0)}{\cancel{p(r)}}\,C(\text{choose } 1|s=0)$

$<$

$\frac{p(s=1)\,p(r|s=1)}{\cancel{p(r)}}\,C(\text{choose } 0|s=1)$

if $C(\text{choose } 0|s=1)$ very high, more likely to choose 1 (and avoid tiger)

summary: 3 main things to consider:

likelihood, prior, cost

Decoding from multiple neurons

More general case: decode continuous stim from multiple neurons

↳ same general idea: $p(\vec{s})$ prior

$p(\vec{r}|\vec{s})$ likelihood

$p(\vec{s}|\vec{r})$ ~~xxxx~~ posterior

$$= \frac{p(\vec{r}|\vec{s})p(\vec{s})}{p(\vec{r})}$$

Max likelihood:

guess $s^* = \arg\max_s p(\vec{r}|\vec{s})$

Max a posteriori:

guess $s^* = \arg\max_s p(\vec{s}|\vec{r}) = \arg\max \frac{p(\vec{r}|\vec{s})p(\vec{s})}{p(\vec{r})}$

$$= \arg\max p(\vec{r}|\vec{s})p(\vec{s})$$

Example: see slides 1–6

Slide 1 notes: "Gaussian" refers only to shape (this is tuning curve, not distribution)

$f_a(s)$ gives <u>mean</u> firing rate as function of stim.

slide 3 notes: need $p(\vec{r}|\vec{s})$ for both ML + MAP

mean #spikes in $T = f_a(s)T$

observed #spikes in $T = r_a T$

slide 4 notes: $\underset{s}{\operatorname{argmax}}\ p(\vec{r}|s) = \operatorname{argmax} \ln(p(\vec{r}|s))$

$$\ln(p(\vec{r}|s)) = \ln\left(\prod_{a=1}^{N} \frac{(f_a(s)T)^{r_a T}}{(r_a T)!} \exp(-f_a(s)T)\right)$$

$$= \sum_{a=1}^{N} \ln\left((f_a(s)T)^{r_a T}\right) - \ln((r_a T)!) - f_a(s)T$$

$$= \sum_{a=1}^{N} r_a T \ln(f_a(s)T) - \underbrace{\sum_{a=1}^{N} \ln((r_a T)!)}_{\text{no dep. on } s} - \underbrace{\sum_{a=1}^{N} f_a(s)T}_{\text{" const.}}$$

$$= \sum_{a=1}^{N} r_a T \ln(f_a(s)T) + C$$

C const. w.r.t s

to find max likelihood:

$$0 = \frac{d}{ds} \ln(p(\vec{r}|s)) = \sum_{a=1}^{N} r_a T \frac{d}{ds} \ln(f_a(s)T)$$

$$= \sum_{a=1}^{N} r_a T \frac{\frac{d}{ds} f_a(s)\cancel{T}}{f_a(s)\cancel{T}}$$

$$f_a(s) = r_{max}\, e^{-\frac{1}{2\sigma_a^2}(s-s_a)^2}$$

$$\longrightarrow \frac{d}{ds} f_a(s) = r_{max}\, e^{-\frac{1}{2\sigma_a^2}(s-s_a)^2} \times \left(-\frac{1}{\cancel{2}\sigma_a^2}\cancel{2}(s-s_a)\right)$$

$$= -\frac{(s-s_a)}{\sigma_a^2} f_a(s)$$

$$= -\sum_{a=1}^{N} \frac{r_a T (s-s_a)}{\sigma_a^2}$$

$$\longrightarrow s \sum_{a=1}^{N} \frac{r_a T}{\sigma_a^2} = \sum_{a=1}^{N} \frac{s_a r_a T}{\sigma_a^2} \longrightarrow s^* = \frac{\sum_{a=1}^{N} s_a r_a / \sigma_a^2}{\sum_{a=1}^{N} r_a / \sigma_a^2}$$

MAP: $s^* = \arg\max_s \ln(p(s|\vec{r}))$

$\ln(p(s|\vec{r})) =$ ~~$\ln(p(s)) \ln(p(\vec{r}))$~~

$\ln(p(\vec{r}|s)\,p(s)/p(\vec{r})) = \ln(p(\vec{r}|s)) + \ln(p(s)) - \ln(p(\vec{r}))$

$$= \sum_{a=1}^{N} r_a T \ln(f_a(s)T) + C + \ln p(s) + D$$

can solve if $p(s)$ is Gaussian w/ mean s_{prior} + var σ^2_{prior}

(see slide 5 for $s|^{n}$)

* often can't find analytical $s|^{n}$ so do numerical optimization

slide 5 notes: ML + MAP are equivalent when prior is constant

can also use $\boxed{\text{Bayesian inference}}$:

instead of $s^* = \arg\max_s p(s|\vec{r})$

use $s^* = \int ds\, p(s|\vec{r})\, s = E_{p(s|\vec{r})}[s]$

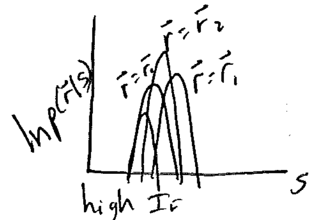
↳ minimizes least squares loss

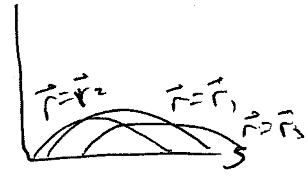
usually hard to compute

$$E_s\left[(s-s^*)^2 \,|\, \vec{r}\right]$$

---

slide 7 notes: averages are over $\vec{r}$'s





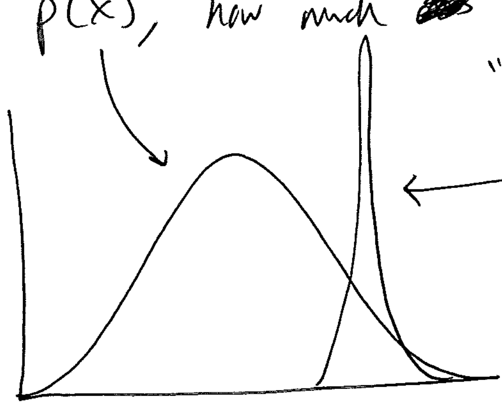
Information Theory

before: talked about estimating stim. from response

now: <u>how much</u> does response tell about stim

↳ info theory provides more general way of quantifying this

Q: given prior $p(x)$, how much "narrower" or "certain" is posterior $p(x|y=y_i)$?

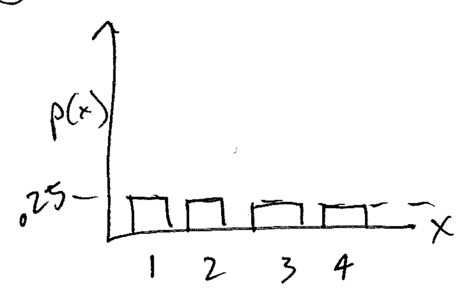


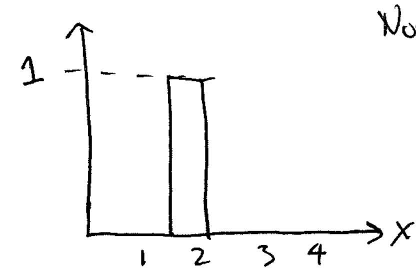
Entropy quantifies "uncertainty" of distribution
(more general than variance + has nice math. properties)

$$\boxed{H[X] = -\sum_x p(x)\log(p(x))}$$

see "A mathematical theory of communication"
Shannon 1948

E.g.





Note: entropy is f^n of <u>entire distribution</u>

$H[X] = -4 \times .25 \times \log .25$
$= \log 4$

$H[X] = -1 \times \log 1 = 0$

(note: $0 \log 0 = 0$)

Mutual Info: $MI(R,S) = H[S] - E_r[H[S|R=r]] = \underbrace{H[R]}_{\text{response entropy}} - \underbrace{E_s[H[R|S=s]]}_{\text{noise entropy}}$

Example:

$s = 0$ (no tiger) or 1 (tiger), $p(s=0) = .9$, $p(s=1) = .1$

$r = 0$ (no spike) or 1 (spike), $p(r=0) = .8$, $p(r=1) = .2$

$p(r=1|s=0) = .1$, $p(r=1|s=1) = .9$

$p(r=0|s=0) = .9$, $p(r=0|s=1) = .1$

no tiger   tiger

How much info does spike/nonspike contain about s?

$$H[R] = -\sum_r p(r)\log p(r) = -[.8\log .8 + .2\log .2] \approx .5$$

$$E_s[H[R|S=s]] = p(s=0)H[R|S=0] + p(s=1)H[R|S=1]$$

$$= .9[-(p(r=0|s=0)\log p(r=0|s=0) + p(r=1|s=0)\log p(r=1|s=0))]$$
$$+ .1[-(p(r=0|s=1)\log p(r=0|s=1) + p(r=1|s=1)\log p(r=1|s=1))]$$

$$= -.9[.9\log .9 + .1\log .1] - .1[.1\log .1 + .9\log .9]$$

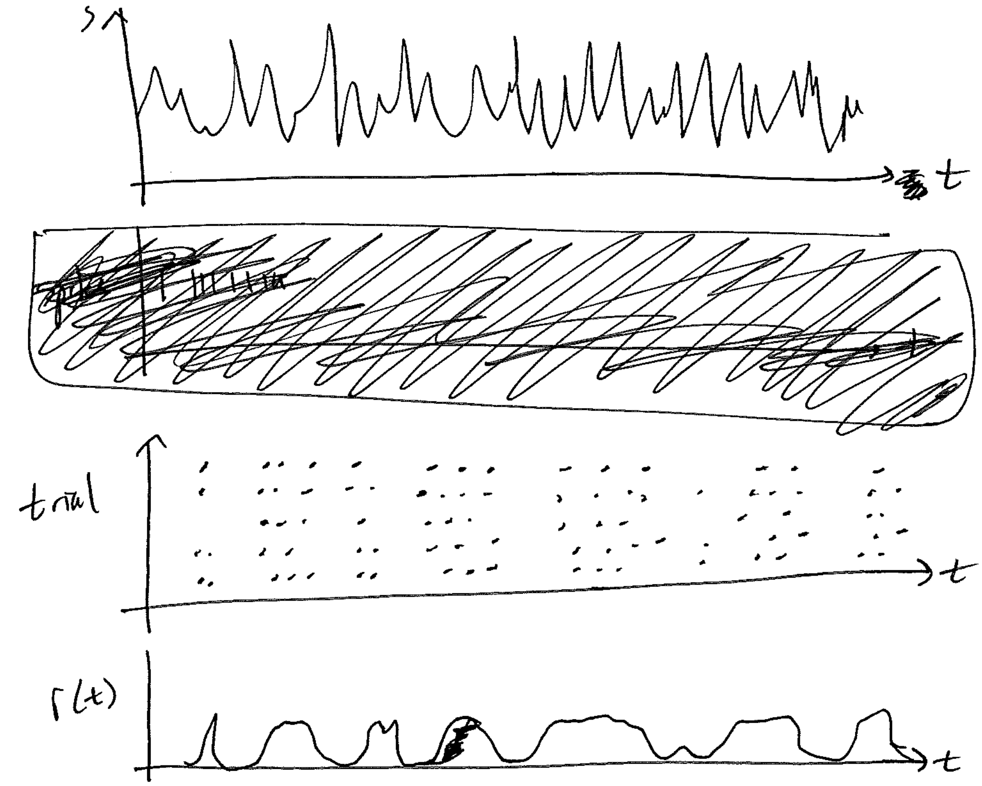
$$= -(.9\log .9 + .1\log .1) \approx .325$$

$$MI(R,S) = H[R] - E_s[H[R|S=s]] \approx .5 - .325 = .175 \text{ bits}$$

What about arbitary stimulus?

see slides 11-12

Context:



avg. rate $= \bar{r}$

$p(r=1) = \bar{r}\Delta t$

$p(r=0) = 1 - \bar{r}\Delta t$

$p(r=1|s) = r(t)\Delta t = p(t)$

$p(r=0|s) = 1 - r(t)\Delta t = 1 - p(t)$

$$MI(R,s) = H(R) - E_{\vec{s}}[H(R|s=\vec{s})]$$

$$= -p\log p - (1-p)\log(1-p)$$

$$= \sum_{\vec{s}} p(\vec{s}) H[R|S=\vec{s}]$$ (hard to calculate!)

But... law of large #s says: $E_{\vec{s}}[H(R|S=\vec{s})] \approx \frac{1}{n}\sum_{i=1}^{n} H[R|S=\vec{s}_i]$

where $\vec{s}_i$ are sampled from $p(\vec{s})$

But $\vec{s}(t)$ are samples from $p(\vec{s})$, so $\longrightarrow \frac{1}{n_t}\sum_{t=1}^{n_t} H[R|S=\vec{s}(t)]$

$$= \frac{1}{n_t}\sum_{t=1}^{n_t} -p(r=0|\vec{s}(t))\log p(r=0|\vec{s}(t)) - p(r=1|\vec{s}(t))\log p(r=1|\vec{s}(t))$$

$$= \frac{1}{n_t}\sum_{t=1}^{n_t} -p(t)\log p(t) - (1-p(t))\log(1-p(t)) \longrightarrow \frac{-1}{T}\int_0^T dt\,[p(t)\log p(t) + (1-p(t))\log(1-p(t))]$$

$\therefore MI(R,S) = H[R] - E_{\vec{s}}(H[R|S=\vec{s}])$

(note: $\bar{r}\Delta t$ is expected spikes in 1 time bin)

$= -p\log p - (1-p)\log(1-p)$

$+ \frac{1}{T}\int_0^T dt\,[p(t)\log p(t) + (1-p(t))\log(1-p(t))]$

Important: doesn't require explicit reference to stimulus, just assumes that $\vec{s}(t)$ was sampled from its prior distribution $p(\vec{s})$

↓

more math

↓

$$MI(R,S) \approx \frac{1}{T}\int_0^T dt\,\frac{r(t)}{\bar{r}}\log\frac{r(t)}{\bar{r}}$$

---

slide 18 notes: dim reduction via STA, covariance analysis, etc.

slide 23 notes: spike trains might contain more info than single spikes treated independently

slide 18 notes: replace $\int_0^T dt$ w/ $\int d^k s\, P(s_1, s_2, \ldots)$

slide 24 notes:

E.g., $\vec{s}(t_1) \rightarrow$

010	trial 1
010	trial 2
011	trial 3
010	trial 4
001	trial 5
011	trial 6

length-3 words: $\vec{S}(t_2) \rightarrow$

000	trial 1
100	trial 2
000	⋮
100	⋮
000	
000	trial 6

⋮

For each stim $s(t)$ get distr. of words $p(w|s(t))$

To calc. $MI(W,S) = H[W] - E_{\vec{s}}[H(W|S=\vec{s})]$

need $p(w)$ — estimate from all words of given length

need $p(w|\vec{s}(t))$ for all t — estimate across trials at time t

use law of large numbers again to go from

$$E_{\vec{s}}[H(W|S=\vec{s})] = \sum_{s} p(\vec{s})\, H[W|S=\vec{s}] \simeq \frac{1}{n_t} \sum_{t} H[W|S=\vec{s}(t)]$$

But as length of words gets longer, harder to estimate $p(w|\vec{s}(t))$.

Therefore, extrapolate:

length	1	2	3	4	…	10
MI(w,S)	85	84	84	83	…	80 …