Role of priors:

Find $z$ by maximizing $P[\text{correct}] = p[+] \beta(z) + p[-](1 - \alpha(z))$
Is there a better test to use than $r$?

The optimal test function is the *likelihood ratio*,

$$l(r) = \frac{p[r|+]}{p[r|-]}.$$  

(Neyman-Pearson lemma)
Penalty for incorrect answer: $L_+, L_-$

For an observation $r$, what is the expected loss?

$$\text{Loss}_- = L_\cdot P[+|r] \quad \text{Loss}_+ = L_\cdot P[-|r]$$

Cut your losses: answer $+$ when $\text{Loss}_+ < \text{Loss}_-$

i.e. when $L_\cdot P[-|r] < L_\cdot P[+|r]$.

Using Bayes’, $P[+|r] = p[r|+]P[+]/p(r)$;
$P[-|r] = p[r|-]P[-]/p(r)$;

$$\Rightarrow l(r) = p[r|+] / p[r|-] > \frac{L_\cdot P[-]}{L_\cdot P[+]}. $$
Decoding from many neurons: population codes

• Population code formulation

• Methods for decoding:
  → population vector
  → Bayesian inference
  → maximum likelihood
  → maximum a posteriori

• Fisher information
Cricket cercal cells

\[
\begin{align*}
\left( \frac{f(s)}{r_{\text{max}}} \right)_a &= \left[ \cos(s - s_a) \right]_+ \\
\left( \frac{\vec{f}(s)}{\vec{r}_{\text{max}}} \right)_a &= [\vec{v} \cdot \vec{c}_a]_+
\end{align*}
\]
Population vector

\[ \vec{v}_{\text{pop}} = \sum_{a=1}^{4} \left( \frac{r}{r_{\text{max}}} \right) \vec{c}_a \]

RMS error in estimate

Theunissen & Miller, 1991
Population coding in M1

Cosine tuning:

\[
\left( \frac{\langle r \rangle - r_0}{r_{\text{max}}} \right)_a = \left( \frac{f(s) - r_0}{r_{\text{max}}} \right)_a = \vec{v} \cdot \vec{c}_a
\]

Pop. vector:

\[
\vec{v}_{\text{pop}} = \sum_{a=1}^{N} \left( \frac{r - r_0}{r_{\text{max}}} \right) \vec{c}_a
\]
Is this the best one can do?

The population vector is neither general nor optimal.

“Optimal”:

make use of all information in the stimulus/response distributions
Bayesian inference

Bayes’ law:

\[ p(s|r) = \frac{p(r|s)p(s)}{p(r)} \]

- likelihood function
- conditional distribution
- prior distribution
- a posteriori distribution
- marginal distribution
Bayesian estimation

Want an estimator $s_{\text{Bayes}}$

Introduce a cost function, $L(s, s_{\text{Bayes}})$; minimize mean cost.

$$
\int ds \ L(s, s_{\text{Bayes}}) p[s | r]
$$

For least squares cost, $L(s, s_{\text{Bayes}}) = (s - s_{\text{Bayes}})^2$.
Let’s calculate the solution.

$$
s_{\text{Bayes}} = \int ds \ p[s | r] s
$$
By Bayes’ law,

\[ p(s|r) = \frac{p(r|s)p(s)}{p(r)} \]

\textit{a posteriori} distribution
Maximum likelihood

Find maximum of $p[r|s]$ over $s$

More generally, probability of the data given the “model”

“Model” = stimulus

assume parametric form for tuning curve
By Bayes’ law,

\[ p(s|r) = \frac{p(r|s)p(s)}{p(r)} \]

a posteriori distribution
MAP and ML

ML: \( s^* \) which maximizes \( p[r|s] \)

MAP: \( s^* \) which maximizes \( p[s|r] \)

Difference is the role of the prior: differ by factor \( p[s]/p[r] \)
Comparison with population vector

\[ \vec{v}_{\text{pop}} = \sum_{a=1}^{4} \left( \frac{r}{r_{\text{max}}} \right)_a \vec{c}_a \]
Decoding an arbitrary continuous stimulus

Many neurons “voting” for an outcome.

Work through a specific example

• assume independence
• assume Poisson firing

Noise model: Poisson distribution

\[ P_T[k] = (\lambda T)^k \exp(-\lambda T)/k! \]
Decoding an arbitrary continuous stimulus

E.g. Gaussian tuning curves

\[ f_a(s) = r_{max} \exp \left( -\frac{1}{2} \left[ \frac{(s - s_a)}{\sigma_a} \right]^2 \right) \]

\[ \sum_{a=1}^{N} f_a(s) \text{ const.} \]

.. what is \( P(r_a|s) \)?
Need to know full $P[r|s]$

Assume Poisson:

$$P[r_a|s] = \frac{(f_a(s)T)^{r_aT}}{(r_aT)!} \exp(-f_a(s)T)$$

Assume independent:

$$P[r|s] = \prod_{a=1}^{N} \frac{(f_a(s)T)^{r_aT}}{(r_aT)!} \exp(-f_a(s)T)$$

Population response of 11 cells with Gaussian tuning curves
Apply ML: maximize $\ln P[r|s]$ with respect to $s$

$$\ln P[r|s] = T \sum_{a=1}^{N} r_a \ln(f_a(s)) + \ldots$$

Set derivative to zero, use sum = constant

$$\sum_{a=1}^{N} r_a \frac{f'(s^*)}{f(s^*)} = 0$$

From Gaussianity of tuning curves,

$$s^* = \frac{\sum r_a s_a / \sigma_a^2}{\sum r_a / \sigma_a^2}$$

If all $\sigma$ same

$$s^* = \frac{\sum r_a s_a}{\sum r_a}$$
Apply MAP: maximise $\ln p[s|r]$ with respect to $s$

$$\ln p[s|r] = \ln P[r|s] + \ln p[s] - \ln P[r]$$

$$\ln p[s|r] = T \sum_{a=1}^{N} r_a \ln(f_a(s)) + \ln p[s] + \ldots$$

Set derivative to zero, use sum = constant

$$\sum_{a=1}^{N} r_a \frac{f'(s^*)}{f(s^*)} + \frac{p'[s]}{p[s]} = 0$$

From Gaussianity of tuning curves,

$$s^* = \frac{T \sum r_a s_a / \sigma_a^2 + s_{\text{prior}} / \sigma_{\text{prior}}^2}{T \sum r_a / \sigma_a^2 + 1 / \sigma_{\text{prior}}^2}$$
Given this data:

**Prior with mean -2, variance 1**

MAP:

**Constant prior**
How good is our estimate?

For stimulus $s$, have estimated $s_{\text{est}}$

**Bias:**

$$b_{\text{est}}(s) = \langle s_{\text{est}} - s \rangle$$

**Variance:**

$$\sigma_{\text{est}}^2(s) = \langle (s_{\text{est}} - \langle s_{\text{est}} \rangle)^2 \rangle$$

**Mean square error:**

$$\langle (s_{\text{est}} - s)^2 \rangle = \langle (s_{\text{est}} - \langle s_{\text{est}} \rangle + b_{\text{est}}(s))^2 \rangle = \sigma_{\text{est}}^2(s) + b_{\text{est}}^2(s).$$

**Cramer-Rao bound:**

$$\sigma_{\text{est}}^2 \geq \frac{(1 + b'_{\text{est}})^2}{I_F(s)}$$

Fisher information

(ML is unbiased: $b = b' = 0$)
Fisher information

\[ I_F(s) = \left\langle -\frac{\partial^2 \ln p[r|s]}{\partial^2 s} \right\rangle = \int d\mathbf{r} p[\mathbf{r}|s] \left( -\frac{\partial^2 \ln p[\mathbf{r}|s]}{\partial^2 s} \right) \]

Alternatively:

\[ I_F(s) = \left\langle \left( \frac{\partial \ln p[\mathbf{r}|s]}{\partial s} \right)^2 \right\rangle = \int d\mathbf{r} p[\mathbf{r}|s] \left( \frac{\partial \ln p[\mathbf{r}|s]}{\partial s} \right)^2 \]

Quantifies local stimulus discriminability
Fisher information for Gaussian tuning curves

For the Gaussian tuning curves w/Poisson statistics:

\[ I_F(s) = \left\langle \left( \frac{d^2 \ln P[r|s]}{ds^2} \right) \right\rangle = T \sum_{a=1}^{N} \langle r_a \rangle \left( \left( \frac{f'_a(s)}{f_a(s)} \right)^2 - \frac{f''_a(s)}{f_a(s)} \right) \]
Are narrow or broad tuning curves better?

\[ I_F = T \sum_{a=1}^{N} \frac{r_{\text{max}}(s - s_a)^2}{\sigma_r^4} \exp \left( -\frac{1}{2} \left( \frac{s - s_a}{\sigma_r} \right)^2 \right) \]

Approximate:

\[ I_F \sim \frac{\sqrt{2\pi} \rho_s \sigma_r r_{\text{max}} T}{\sigma_r^2}. \]

Thus,

\[ I_F \sim 1/\sigma_r \quad \rightarrow \quad \text{Narrow tuning curves are better} \]

But not in higher dimensions!

\[ I_F \sim (2\pi)^{D/2} D \rho_s \sigma_r^{D-2} r_{\text{max}} T \]

..what happens in 2D?
Fisher information and discrimination

Recall $d' = \text{mean difference/standard deviation}$

Can also decode and discriminate using decoded values.

Trying to discriminate $s$ and $s+\Delta s$:

Difference in ML estimate is $\Delta s$ (unbiased) variance in estimate is $1/I_F(s)$.

$\rightarrow \quad d' = \Delta s \sqrt{I_F(s)}$
Limitations of these approaches

• Tuning curve/mean firing rate

• Correlations in the population
The importance of correlation

Shadlen and Newsome, ‘98
The importance of correlation
The importance of correlation
Entropy and Shannon information

Model-based vs model free