

Last time:

- 1) Population coding
- 2) Bias & variance of estimators
- 3) Fisher information

HW5 # updated

Recall Fisher information:

$$\begin{aligned} I_F(s) &= E \left[\left(\frac{\partial}{\partial s} \log p(r|s) \right)^2 \right]_{p(r|s)} \\ &= -E \left[\frac{\partial^2}{\partial s^2} \log p(r|s) \right]_{p(r|s)} \end{aligned}$$

Properties: 1) Local (dependent on value of s)

2) If r_i, r_j independent, $I_F(s) = I_F^i(s) + I_F^j(s)$

$$\log p(\underline{r}|s) = \log p(r_i|s) + \log p(r_j|s)$$

3) More generally, $I_F^{X,Y} = I_F^X + I_F^Y$

5) Related

to variance of unbiased estimator

4) Dependent on stimulus parameterization:

$$\text{If } u = f(s), I_F(s) = I_F(u) \left(\frac{du}{ds} \right)^2 \quad (\text{Cramér-Rao bound, later}).$$

Equivalence of 2 definitions:

$$\frac{\partial^2}{\partial s^2} \log p(\underline{r}|s) = \frac{\partial}{\partial s} \left[\frac{\partial}{\partial s} \log p(\underline{r}|s) \right]$$

$$= \frac{\partial}{\partial s} \left[\frac{1}{p(\underline{r}|s)} \frac{\partial}{\partial s} p(\underline{r}|s) \right]$$

$$= \underbrace{\frac{1}{p(\underline{r}|s)} \frac{\partial^2}{\partial s^2} p(\underline{r}|s)}_{E[\cdot]} - \underbrace{\left(\frac{\frac{\partial}{\partial s} p(\underline{r}|s)}{p(\underline{r}|s)} \right)^2}_{= \left[\frac{\partial}{\partial s} \log p(\underline{r}|s) \right]^2}$$

$$E[\cdot] = \int d\underline{r} \frac{\partial^2}{\partial s^2} p(\underline{r}|s)$$

$$= \frac{\partial^2}{\partial s^2} \int d\underline{r} p(\underline{r}|s) = 0$$

$$= \left[\frac{\partial}{\partial s} \log p(\underline{r}|s) \right]^2$$

$$\text{So } I_F(s) = E \left[\left(\frac{\partial}{\partial s} \log p(\underline{r}|s) \right)^2 \right]_{p(\underline{r}|s)}$$

Cramér-Rao bound: If $\hat{\theta}$ is an unbiased estimator of a parameter θ of a distribution $p(r|\theta)$, then

$$\text{Var}(\hat{\theta}|\theta) \geq \frac{1}{I_F(\theta)}$$

$$\text{Recall unbiased} \Rightarrow E[\hat{\theta}|\theta] - \theta = 0.$$

Proof: Score $A = \frac{\partial}{\partial \theta} \log p(r|\theta)$. From last time, $E[A] = 0$.

$$\text{Also let } B = \hat{\theta} - \theta \quad E[B] = 0$$

Use Cauchy-Schwartz inequality:

$$E[AB]^2 \leq E[A^2] E[B^2]$$

$$E[A^2] = I_F(\theta)$$

$$E[B^2] = \text{Var}(\hat{\theta}|\theta)$$

$$\begin{aligned} E[AB] &= \int dr \left[(\hat{\theta} - \theta) \frac{\partial}{\partial \theta} \log p(r|\theta) \right] p(r|\theta) \\ &= \underbrace{\int dr \hat{\theta} \frac{\partial}{\partial \theta} \log p(r|\theta) p(r|\theta)}_{(1)} - \underbrace{\int dr \theta \frac{\partial}{\partial \theta} \log p(r|\theta) p(r|\theta)}_{(2)} \end{aligned}$$

$$(1) = \frac{d}{d\theta} \int d\underline{r} \underbrace{\hat{\theta}}_{\hat{\theta}} p(\underline{r}|\theta) = \frac{d}{d\theta} (1) = 1 \quad (\text{note } \hat{\theta} \text{ is a function of } \underline{r}, \text{ not } \theta \text{ explicitly})$$

$$(2) = \theta \int d\underline{r} \frac{d}{d\theta} \log p(\underline{r}|\theta) \hat{\theta} p(\underline{r}|\theta) = \theta E\left[\frac{d}{d\theta} \log p(\underline{r}|\theta)\right] = \theta E[\text{Score}] = 0$$

$$\Rightarrow 1 \leq \mathcal{I}_F(\theta) \times \text{Var}(\hat{\theta}|\theta)$$

$\hookrightarrow = 0$ from last time.

$$\Rightarrow \text{Var}(\hat{\theta}|\theta) \geq 1/\mathcal{I}_F(\theta)$$

Generalizations:

1) Biased estimators:

If $\hat{\theta}$ has bias $b(\hat{\theta}|\theta)$, then

$$\text{Var}(\hat{\theta}|\theta) \geq \frac{[1 + b'(\hat{\theta}|\theta)]^2}{\mathcal{I}_F(\theta)}$$

$$b' = \frac{d}{d\theta} b(\hat{\theta}|\theta)$$

2) Multivariate case

Fisher information matrix:

$$[\mathcal{I}_F(\underline{s})]_{ij} = E\left[\left(\frac{d}{ds_i} \log p(\underline{r}|\underline{s})\right) \left(\frac{d}{ds_j} \log p(\underline{r}|\underline{s})\right) \mid \underline{s}\right]$$

$$= -E\left[\frac{d}{ds_i ds_j} \log p(\underline{r}|\underline{s}) \mid \underline{s}\right]$$

$$\underline{a}^T \text{Cov}(\underline{\hat{S}} | \underline{s}) \underline{a} \geq \underline{a}^T \left[\underline{I}_F(\underline{s}) \right]^{-1} \underline{a} \quad \forall \underline{a}$$

(diagonal elements $\Rightarrow$ univariate case)

Covariance Matrices

$$[\Sigma(\underline{x})]_{ij} = \text{Cov}(x_i, x_j) = E[(x_i - E[x_i])(x_j - E[x_j])]$$

$$\Sigma = \begin{pmatrix} \text{Var}(x_1) & \text{Cov}(x_1, x_2) & \dots \\ \text{Cov}(x_2, x_1) & \text{Var}(x_2) & \\ \vdots & & \ddots \\ & & & \text{Var}(x_n) \end{pmatrix}$$

Properties: 1) Symmetric

2) Positive-semidefinite: $\underline{a}^T \Sigma \underline{a} \geq 0$

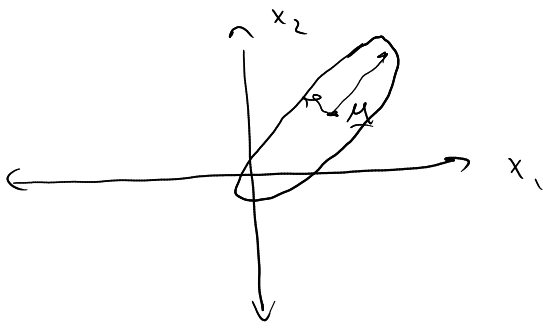
3) E-vals, e-vects related to PCA. Note

Σ, Σ^{-1} have same e-vects (see with SVD).

Multivariate normal distribution:

$$E[\underline{x}] = \underline{\mu}, \quad \text{Cov}(\underline{x}) = \Sigma \quad \begin{matrix} \underline{x} \sim \mathcal{N}(\underline{\mu}, \Sigma) \\ \underline{x} \in \mathbb{R}^N \end{matrix}$$

$$p(\underline{x}) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}} \exp\left(-\frac{1}{2}(\underline{x} - \underline{\mu})^T \Sigma^{-1}(\underline{x} - \underline{\mu})\right)$$



Σ diagonal: independent.

Linear Fisher information (Moreno-Bote et al., Nat Neurosci 2014)

Suppose N neurons w/ tuning curves $E[r_i | s] = \mu_i(s)$,
 $i = 1 \dots N$. Let $\underline{\mu}(s) = \begin{bmatrix} \mu_1(s) \\ \mu_2(s) \\ \vdots \\ \mu_N(s) \end{bmatrix}$

and $\underline{\mu}'(s) = \frac{\partial}{\partial s} \underline{\mu}(s)$. (tuning curve slope).

Also assume that $r_i(s) \sim \mathcal{N}(\underline{\mu}(s), \Sigma(s))$.

Note: $\Sigma(s)$ noise covariance matrix. Correlation:

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii} \Sigma_{jj}}}$$

Can find: $I_F(\) = (\underline{\mu}')^T \Sigma^{-1} \underline{\mu}' +$

$$\frac{1}{2} \text{Tr}[(\Sigma' \Sigma^{-1})^2]$$

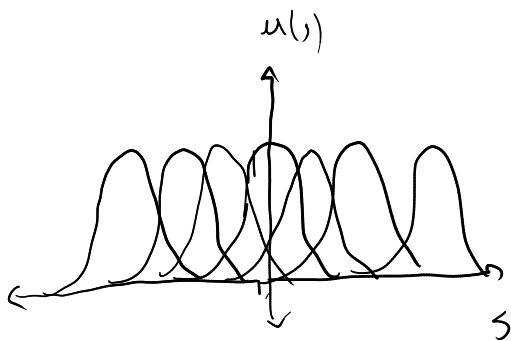
$\Sigma' = \frac{\partial}{\partial s} \Sigma(s)$: Stimulus-dependent noise correlations.

If we consider a ^{"locally optimal"} linear estimator $\hat{s} = \sum_i w_i r_i$,

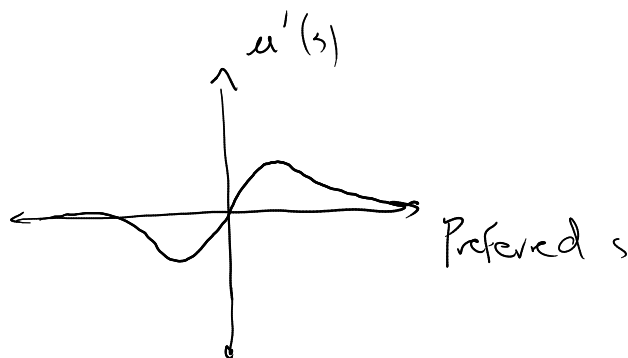
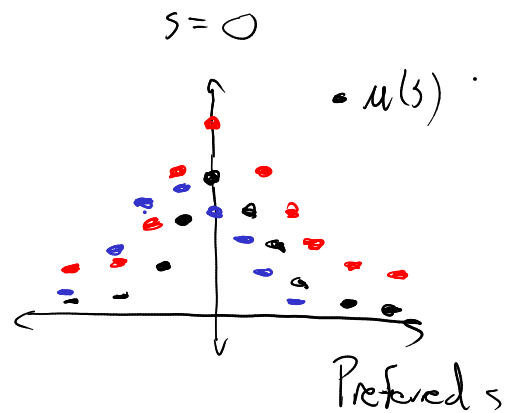
$$\text{Var}(\hat{\theta}|\theta) \geq \frac{1}{I_{\text{LOLE}}(s)}, \text{ where}$$

$$I_{\text{LOLE}}(s) = (\underline{\mu}')^T \Sigma^{-1} \underline{\mu}'$$

Also valid if $p(\underline{r}|s)$ is of exponential family w/ linear sufficient statistics.



$\Rightarrow$



Consequences:

1) Correlations in the direction of $(\underline{\mu}')^T \underline{\mu}'$ limit information as $N \rightarrow \infty$. "Differential correlations".
Identical tuning curves $\Rightarrow$ pos. mean correlations.

If $\Sigma = \Sigma_0 + \varepsilon (\underline{\mu}')^T \underline{\mu}'$, where $I_0(\cdot) \rightarrow \infty$
as $N \rightarrow \infty$ if $\Sigma = \Sigma_0$, then

$$I = \frac{I_0}{1 + \varepsilon I_0}. \quad \text{As } N \rightarrow \infty, \quad I \rightarrow 1/\varepsilon$$

2) Differential correlations may be hidden.
Instead, use decoder.