Bias & variance:

Bias of $\hat{s}$: $b(s) = E[\hat{s}\,|s] - s$

(defined for any parameter Θ of a statistical model)

$$\mathrm{Var}(\hat{s}|s) = E\left[(\hat{s} - E[\hat{s}|s])^2\,|s\right]$$

$$E\left[(\hat{s}-s)^2\right] = E\left[\left(\hat{s} - E[\hat{s}|s] + b(s)\right)^2\,|s\right]$$

$$= \underbrace{\mathrm{Var}(\hat{s}|s)}_{\text{variance}} + \underbrace{b^2(s)}_{\text{bias}}$$

$b(s) = 0$: unbiased estimator

Fisher information (scalar case):

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

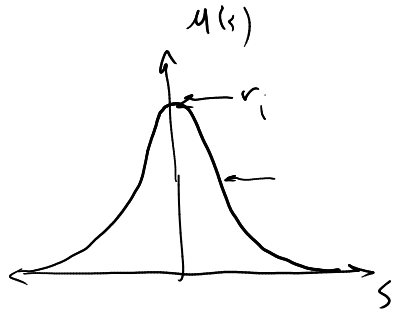
$$= -\int d\underline{r}\; p(\underline{r}|s)\; \frac{\partial^2 \log p(\underline{r}|s)}{\partial s^2}$$

or

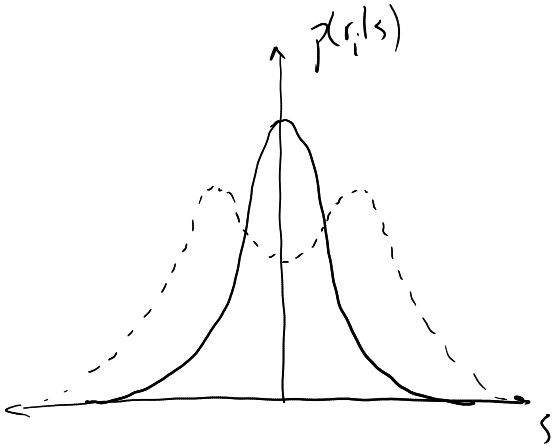
$$E\left[\left(\frac{\partial}{\partial s}\log p(\underline{r}|s)\right)^2\right]_{p(\underline{r}|s)} \qquad (2)$$

Example: $\lambda_i = u(s)$ $\quad r_i \sim \text{Pois}(\lambda_i)$

↑ tuning curve



$$p(r_i|s) = \frac{u(s)^{r_i} e^{-u(s)}}{r_i!}$$



$$\log p(r_i|s) = r_i \log u(s) - u(s) - \log r_i!$$

$$\frac{\partial}{\partial s} \log p(r_i|s) = r_i \frac{u'(s)}{u(s)} - u'(s) = u'(s)\left[\frac{r_i}{u(s)} - 1\right]$$

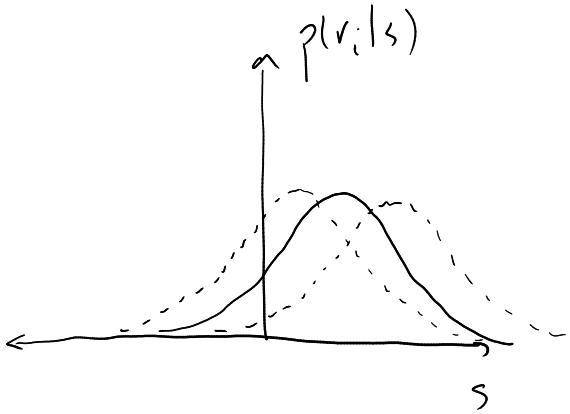
$$E_{p(r_i|s)}\left[(\cdot)^2\right] = \left(u'(s)\right)^2 \underbrace{E\left[\frac{r_i^2}{u^2(s)} - \frac{2r_i}{u(s)} + 1\right]}$$

$E[r_i^2] = u(s)(1+u(s))$

$E[r_i] = u(s)$

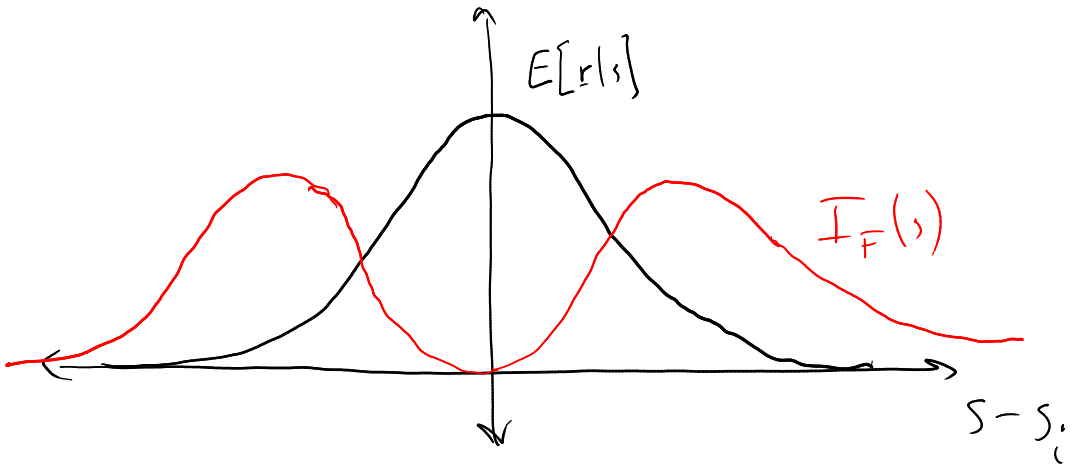
$$\frac{1}{u(s)} + 1 - 2 + 1$$

$$= \frac{\left(u'(s)\right)^2}{u(s)}$$



$$\lambda_i = r_{max} \exp\left(\frac{-(s_i - s)^2}{2\sigma_i^2} \right) \qquad r_i \sim Pois(\lambda_i)$$

$$I_F(s) = \frac{r_{max} (s_i - s)^2}{\sigma_i^4} \exp\left(\frac{-(s_i - s)^2}{2\sigma_i^2} \right)$$



Intuition (using def 1): Expected curvature of log-likelihood fn.

Intuition (using def. 2): "Score": $\frac{d}{ds} \log p(\underline{r}|s)$

How much does log-likelihood of observing $\underline{r}$ change when s is varied?

$$E\left[\frac{d}{ds} \log p(\underline{r}|s) \right]_{p(\underline{r}|s)} = \int d\underline{r}\, \cancel{p(\underline{r}|s)} \cdot \frac{\frac{d}{ds} p(\underline{r}|s)}{\cancel{p(\underline{r}|s)}}$$

$$= \frac{d}{ds} \int d\underline{r}\, p(\underline{r}|s) = \frac{d}{ds}(1) = 0.$$

So $\quad Var(\text{score}) = E[\text{score}^2] = I_F(s)$

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, $I_F^{X,Y} = I_F^X + I_F^{Y|X}$

4) Dependent on stimulus parameterization:

If $u = g(s)$, $I_F(s) = I_F(u)\left(\frac{du}{ds}\right)^2$

5) Related to variance of unbiased estimator (Cramér-Rao bound, later).

Equivalence of 2 definitions:

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

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

$$= \underbrace{\frac{1}{p(\underline{r}|s)}\frac{\partial^2}{\partial s^2}p(\underline{r}|s)} - \underbrace{\left(\frac{\frac{\partial}{\partial s}p(\underline{r}|s)}{p(\underline{r}|s)}\right)^2}$$

$$E[\cdot] = \int d\underline{r}\,\frac{\partial^2}{\partial s^2}p(\underline{r}|s) = \frac{\partial^2}{\partial s^2}\int d\underline{r}\,p(\underline{r}|s) = 0$$

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

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

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

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

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

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

$E[A] = 0$.

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

Use Cauchy-Schwartz inequality:

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

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

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

$$E[AB] = \int d\underline{r} \left[(\hat{\theta} - \theta) \frac{\partial}{\partial \theta} \log p(\underline{r}|\theta) \right] p(\underline{r}|\theta)$$

$$= \underbrace{\int d\underline{r}\, \hat{\theta} \frac{\partial}{\partial \theta} \log p(\underline{r}|\theta)\, p(\underline{r}|\theta)}_{(1)} - \underbrace{\int d\underline{r}\, \theta \frac{\partial}{\partial \theta} \log p(\underline{r}|\theta)\, p(\underline{r}|\theta)}_{(2)}$$

$$(1) = \frac{\partial}{\partial\theta}\int d\underline{r}\;\hat{\theta}\;p(\underline{r}|\theta) = \frac{\partial}{\partial\theta}(\theta) = 1$$

(note $\hat{\theta}$ is a function of $\underline{r}$, not θ explicitly)

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

↳ $=0$ from last time.

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

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

Generalizations:

1) Biased estimators:

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

$$Var(\hat{\theta}|\theta) \geq \frac{[1+b'(\hat{\theta}|\theta)]^2}{\underline{I}_F(\theta)} \qquad b' = \frac{\partial}{\partial\theta}b(\hat{\theta}|\theta)$$

2) Multivariate case

Fisher information matrix:

$$[\underline{I}_F(\underline{s})]_{ij} = E\left[\left(\frac{\partial}{\partial s_i}\log p(\underline{r}|\underline{s})\right)\left(\frac{\partial}{\partial s_j}\log p(\underline{r}|\underline{s})\right)\middle|\,\underline{s}\right]$$

$$= -E\left[\frac{\partial}{\partial s_i \partial s_j}\log p(\underline{r}|\underline{s})\middle|\,\underline{s}\right]$$

$$\underline{a}^T \mathrm{Cov}(\hat{\underline{s}} \mid \underline{s})\, \underline{a} \geq \underline{a}^T \left[I_F(\underline{s}) \right]^{-1} \underline{a} \quad \forall\, \underline{a}$$

(diagonal elements $\Rightarrow$ univariate case)

Covariance Matrices

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

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

Properties: 1) Symmetric

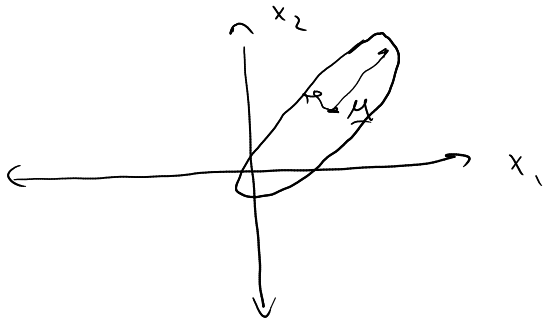
2) Positive-semidefinite: $\underline{a}^T \Sigma\, \underline{a} \geq 0$

3) E-vals, e-vecs related to PCA. Note Σ, Σ^{-1} have same e-vecs (see with SVD).

Multivariate normal distribution:

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

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



Σ diagonal: independent.

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

Suppose N neurons w/ tuning curves $E[r_i|s] = \mu_i(s)$, $i = 1 \ldots N$. 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 $\underline{r}_i(s) \sim N(\underline{\mu}(s), \Sigma(s))$.

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

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

Can find: $I_F(\ ) = (\underline{\mu}')^T \Sigma^{-1} \underline{\mu}' + \frac{1}{2} \mathrm{Tr}\left[\left(\Sigma' \Sigma^{-1}\right)^2\right]$

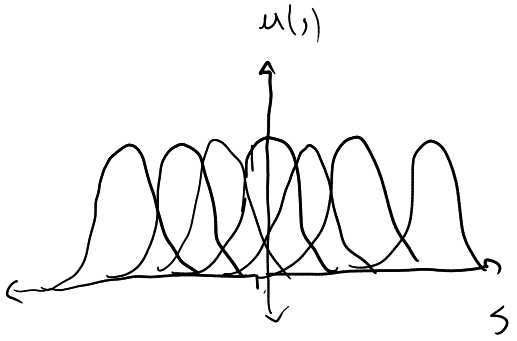
$\Sigma' = \frac{\partial}{\partial s}\Sigma(s)$: Stimulus-dependent noise correlations.

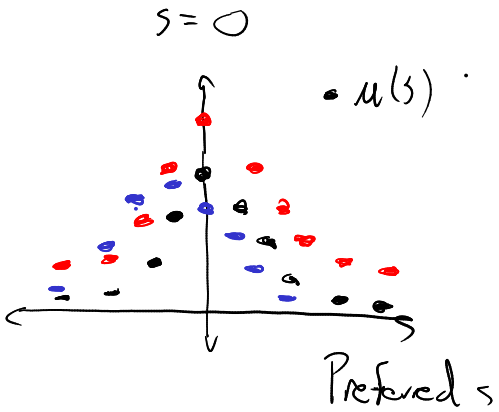
If we consider a "locally optimal" linear estimator $\hat{s} = \sum_i w_i r_i$,

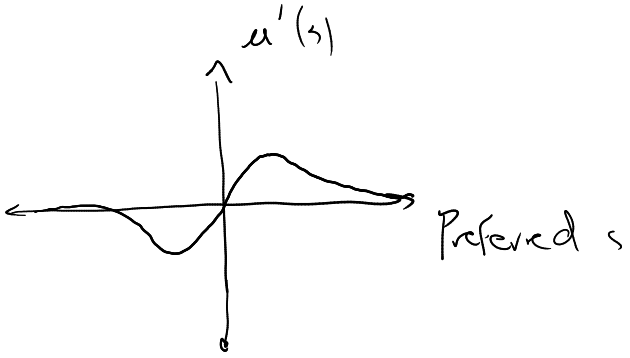
$$\mathrm{Var}(\hat{\theta}|\theta) \geq \frac{1}{I_{LOLE}(s)}, \text{ where}$$

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

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







Consequences:

1) Correlations in the direction of $(\underline{\mu}')^T \underline{\mu}'$ limit information as $N \to \infty$. "Differential correlations"

Identical tuning curves $\Rightarrow$ pos. mean correlations.

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

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

2) Differential correlations may be hidden.

Instead, use decoder.