

# Vision Notes

Contains personal notes on *Fisher Information* and other related/necessary concepts.

## Fisher Information (FI)

From the Simoncelli lecture, the Fisher Information  $I(s)$  of a given stimulus  $s$  is the second-order expansion of the expected negative log likelihood.

### Definition

$$I(s) = -\mathbb{E} \left[ \frac{\partial^2 \log p(r|s)}{\partial s^2} \right] \quad (1)$$

### Intuition

It is handy to think about Fisher information through the lense of perception studies, which use this technique a lot. For our scenario, we can consider an experiment where we wish to find thresholds around human perception of brightness. For example, how much must we increase the brightness value of a light in order for us to perceive the change? The light is the **stimulus** - it is the input from the outside world that we receive. Given this stimulus, our eyes, brain, etc will produce a **response**. There is not a one-to-one mapping of stimulus to response. Due to biological noise, a given stimulus value will always correspond with a **response distribution**. In other words, a given stimulus could produce a range of response values within our internals, but these values will follow some distribution with a certain mean, variance, etc. It is then helpful to think of this function between stimuli (on the x-axis) and response (on the y-axis) with an subsequent function  $f(s)$  applied. For example, we will use  $f$  as the mean of the response corresponding with the stimulus  $s$ .

We are not particularly interested in the response itself however - that part is internal to the brain, and hard to study. We care about the *estimate* of the stimulus that we make, which is based on the internal response. To summarize, our brain follows the following process:

1. Eyes receive the stimulus and transmit the signal to the brain.
2. Brain creates a response from this stimulus signal. This is a stochastic process - the same stimulus value could produce many response values. Even more importantly, **multiple stimuli values could produce the same response value**.
3. Based on the shape, mean, etc of the response, the brain must make an estimation of what stimulus value is most likely to have produced that response.
  - This can be tricky! We note above that different stimulus values might produce the same response value in the brain. The challenge is finding how different these stimulus values have to be in order for us to **discriminate** between them reliably (ie, with some probability).

**Side Note:** I can think of two different ways to increase the odds of a subject discriminating between two stimuli.

1. Increasing the “distance” between the stimuli. Responses to stimuli are distributions with some mean value.
2. Increasing the number of samples taken from the response distribution. This can come from having a longer exposure to the stimulus - for example, viewing a light for a longer period of time allows our brain to take more samples from the response distribution corresponding with the brightness level of the light.

- The idea of having more exposure is pretty intuitive from a statistical standpoint. It is equivalent to increasing the sample size, which we know will produce more accurate estimations of the mean, variance, etc of a distribution.

The point of the above paragraph is to say that the brain has to rely on the mean, variance, etc of

Note that this second derivative expression gives us the precision of an estimator.

### Alternative Definitions

*Probability and Statistics* describes the FI as, “one property of a distribution that can be used to measure how much information one is likely to obtain from a random variable or a random sample.” It receives a primary definition in a slightly different (more generalized) format:

$$I(\theta) = \mathbb{E}_{\theta} \{ [\lambda'(X|\theta)]^2 \} \quad (2)$$

where  $\lambda(x|\theta)$  is the log probability of a random variable  $X$ , that is

$$\lambda(x|\theta) = \log f(x|\theta) \quad (3)$$

for  $X$ 's pdf  $f(x|\theta)$  with parameter  $\theta$ . It is further assumed that  $\lambda(x|\theta)$  is twice differentiable with respect to  $\theta$ . This textbook definition is completely equivalent to the one used in the NYU lecture. The differences in formatting comes from the fact that the NYU lecture is focused on psychophysics and perception, while the textbook looks to generalize the FI to any application.

If  $f$  is indeed a pdf, then the FI can be calculated as

$$I(\theta) = \int_S [\lambda'(x|\theta)]^2 f(x|\theta) dx. \quad (4)$$

$X$  takes on values in the sample space  $S$ , which is why we take the integral over  $S$  when calculating the expectation. Note that  $f$  can also be pmf (when  $X$  is discrete). In this case, we just take the sum over all  $x$  in  $S$  when calculating the expectation.

We can also express the FI in these two ways, which are equivalent to the above definition:

$$I(\theta) = -\mathbb{E}_{\theta} [\lambda''(X|\theta)] \quad (5)$$

$$I(\theta) = \text{Var}_{\theta} [\lambda'(X|\theta)] \quad (6)$$

**Proof.** For clarity, here are all of the conditions/assumptions required for the previous definitions and this proof:

- $X$  is a random variable, with a distribution that depends on a parameter  $\theta$  that takes values in an open interval  $\Omega$  of the real line.
- $f(x|\theta)$  is the probability density function of  $X$ . The same definitions and proof hold if  $f$  is a mass function, but for simplicity we'll show just the continuous case.
- The set of  $x \in X$  such that  $f(x|\theta) > 0$  is the same for all  $\theta$ .
- $\lambda(x|\theta) = \log f(x|\theta)$  is twice differentiable as a function of  $\theta$ .
- The first and second derivatives of  $\int_S f(x|\theta) dx$  with respect to  $\theta$  can be calculated by reversing the order of integration and differentiation (ie. differentiating  $f$ , then taking the integral of that expression).

We know that  $\int_S f(x|\theta) dx = 1$  for all  $\theta \in \Omega$ . This means that if the integral expression on the left-hand side of the equation were written as a function of  $\theta$ , it would be a constant function

of 1. Therefore, the result of differentiating this expression with respect to  $\theta$  is 0. We will assume that we can take the derivative inside the integral sign and receive

$$\int_S f'(x|\theta)dx = 0 \text{ for } \theta \in \Omega. \quad (7)$$

We also made the assumption that we can take the second derivative with respect to  $\theta$  inside of the integral, so we receive

$$\int_S f''(x|\theta)dx = 0 \text{ for } \theta \in \Omega. \quad (8)$$

The derivative of  $\lambda$  is

$$\begin{aligned} \frac{d}{dx}(\lambda(x|\theta)) &= \frac{d}{dx}(\log(f(x|\theta))) \\ &= \frac{1}{f(x|\theta)} \cdot \frac{d}{dx}(f(x|\theta)) \\ &= \frac{f'(x|\theta)}{f(x|\theta)}. \end{aligned} \quad (9)$$

Knowing this, we can write the following expectation as

$$\begin{aligned} \mathbb{E}_\theta[\lambda'(X|\theta)] &= \int_S \lambda'(x|\theta)f(x|\theta)dx \\ &= \int_S \left( \frac{f'(x|\theta)}{f(x|\theta)} \right) \cdot f(x|\theta)dx \\ &= \int_S f'(x|\theta)dx. \end{aligned} \quad (10)$$

Therefore it follows from Eq. 7 that

$$\mathbb{E}_\theta[\lambda'(X|\theta)] = 0. \quad (11)$$

Note that one formula for variance is  $\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$  where  $\mu$  is the mean of  $Y$ . We want to show that the expression  $\text{Var}_\theta[\lambda'(X|\theta)]$  is equivalent to the fisher information.

$$\begin{aligned} \text{Var}_\theta[\lambda'(X|\theta)] &= \mathbb{E}\{[\lambda'(X|\theta)]^2\} - \mathbb{E}[\lambda'(X|\theta)]^2 \\ &= \mathbb{E}\{[\lambda'(X|\theta)]^2\} - (0)^2 \quad \text{Eq. 11} \\ &= \mathbb{E}\{[\lambda'(X|\theta)]^2\}. \end{aligned} \quad (12)$$

Therefore Eq. 6 is a valid definition of the FI. To show that Eq. 5 is one as well, we need to take the second derivative of  $\lambda(X|\theta)$ . We'll make use of the quotient rule, listed here for reference:

$$\frac{d}{dx} \left( \frac{f(x)}{g(x)} \right) = \frac{g(x) \cdot f'(x) - f(x) \cdot g'(x)}{g(x)^2} \quad (13)$$

The second derivative of  $\lambda$  is then

$$\begin{aligned}
\lambda''(x|\theta) &= \frac{d}{dx}(\lambda'(x|\theta)) \\
&= \frac{d}{dx} \left( \frac{f'(x|\theta)}{f(x|\theta)} \right) \\
&= \frac{f(x|\theta) \cdot f''(x|\theta) - f'(x|\theta) \cdot f'(x|\theta)}{[f(x|\theta)]^2} \\
&= \frac{f(x|\theta) \cdot f''(x|\theta)}{[f(x|\theta)]^2} - \frac{[f'(x|\theta)]^2}{[f(x|\theta)]^2} \\
&= \frac{f''(x|\theta)}{f(x|\theta)} - [\lambda'(x|\theta)]^2.
\end{aligned} \tag{14}$$

We now want to show that the expression  $-\mathbb{E}_{\theta[\lambda''(X|\theta)]}$  equals our original definition of the Fisher information, Eq. 2.

$$\begin{aligned}
-\mathbb{E}_{\theta}[\lambda''(X|\theta)] &= -\mathbb{E}_{\theta} \left[ \frac{f''(x|\theta)}{f(x|\theta)} - [\lambda'(x|\theta)]^2 \right] \\
&= -\mathbb{E}_{\theta} \left[ \frac{f''(x|\theta)}{f(x|\theta)} \right] - \mathbb{E}_{\theta}[-\lambda'(x|\theta)]^2 \\
&= \int_S \left[ \frac{f''(x|\theta)}{f(x|\theta)} \cdot f(x|\theta) dx \right] + \mathbb{E}_{\theta}[\lambda'(x|\theta)]^2 \\
&= \int_S [f''(x|\theta) dx] + \mathbb{E}_{\theta}[\lambda'(x|\theta)]^2 \\
&= (0) + \mathbb{E}_{\theta}[\lambda'(x|\theta)]^2 \tag{Eq. 8} \\
&= \mathbb{E}_{\theta}[\lambda'(x|\theta)]^2
\end{aligned} \tag{15}$$

Therefore Eq. 5 is also a valid definition of the FI. ■

### Intuition Behind Alt. Definitions

The first thing to note is that Eq. 5 is the same exact expression as the one taught in the NYU lecture. Often, it is easier to determine and/or compute the FI using this equation compared to the other two listed here.

### Tools

Fisher information allows us to place a bound on the “precision” of unbiased estimators. This is known as the **Cramer-Rao** bound:

$$\sigma^2(s) \geq \frac{1}{I(s)} \tag{16}$$

When dealing with perception, this property is useful for providing a bound on discrimination  $D$ :

$$D(s) \leq \sqrt{I(s)} \tag{17}$$

## Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations

### Overview

## Notation & Equations

$N$

- the population size, describes a set of  $N$  neurons

$n$

- index into the current neuron (eg: “Assume the number of spikes emitted in a given time interval by the  $n$ th neuron ...”)

$h_n(s)$

- the tuning function of the  $n$ th neuron

$R$

- the total expected spike rate of the neuron population

**Assuming** the number of spikes emitted in a given time interval by the  $n$ th neuron is a sample from an independent **Poisson** process, with a **mean rate** determined by its tuning function  $h_n(s)$ , the *probability* distribution of the population response is written as:

$$p(*r*|s) = \prod_{n=1}^N \frac{h_n(s)^{r_n} e^{-h_n(s)}}{r_n!} \quad (18)$$