

Score, Fisher Information and MMSE

References:

1. Prof. Krishna Narayanan, Texas A&M Univ.,
Personal Communication, 2025.
2. L. Brown, Admissible estimators, recurrent diffusions, and
insoluble boundary value problems. The Annals of Mathematical
Statistics, Vol. 42, No. 3, 1971.
3. W. Cao et. al., Non-parametric estimation of Fisher
Information and its applications, arXiv 2005.03622, 2020
4. D. Guo, S. Shamai and S. Verdú, Mutual Information
and minimum mean-square error in Gaussian channels,
IEEE Trans. IT, 2007.

When $x_0, x_\sigma \in \mathbb{R}^d$, we have:

$$x_\sigma = x_0 + \sigma Z,$$

$$Z \sim N(0, I_d)$$

$\downarrow$
 $\sim p_0(\cdot)$

$\downarrow$
 $\sim p_0(\cdot)$

$$x_0 \perp\!\!\!\perp Z$$

$$S_\sigma(x_\sigma) = \nabla_{x_\sigma} \ln p_\sigma(x_\sigma)$$

$\downarrow$
score at σ -noise level.

$$\text{MMSE}(x_\sigma; \sigma) = E[x_0 | x_\sigma = x_\sigma]$$

$\downarrow$ observation

$\downarrow$ noise level

minimum
mean-square
estimate

(in information theory,
 $\text{SNR} = \frac{1}{\sigma^2}$)

$$E[x_0 | x_\sigma = x_\sigma] = x_\sigma + \sigma^2 \nabla \ln p_\sigma(x_\sigma).$$

(Tweedie's Formula)

$$\boxed{\text{MMSE}(x_\sigma; \sigma) = x_\sigma + \sigma^2 S_\sigma(x_\sigma).}$$

Recall from Note 4 (Markov Processes), that the Fisher Information is defined as:

$$\text{FI}(p_\sigma) = E_{x_0 \sim p_\sigma} [\|\nabla_x \ln p_\sigma(x_0)\|^2]$$

Fisher Information

$$\boxed{\text{FI}(p_\sigma) = E_{x_0 \sim p_\sigma} [\|S_\sigma(x_0)\|^2]}$$

(we are overloading notation p_σ to represent both the density function as well as the measure $p_\sigma(x) dx$)

To complete this, we next show that MMSE and FI are directly related.

$$\begin{aligned} \text{Let } \mathcal{E}(p_\sigma; \sigma) &= E_{(x_0, x_\sigma)} [\|x_0 - E[x_0 | x_\sigma]\|^2] \\ &\quad \text{MMSE Error} \\ &= E_{\substack{x_0 \sim p_0 \\ z \sim N(0, I)}} [\|x_0 - E[x_0 | x_0 + \sigma z]\|^2] \end{aligned}$$

Then, $\boxed{\mathcal{E}(p_\sigma; \sigma) = \sigma^2 d - \sigma^4 \text{FI}(p_\sigma)}$ $\rightarrow \textcircled{*}$

Proof of (*): Recall: $E[x_0 | x_\sigma = x_\sigma] = x_\sigma + \sigma^2 S_\sigma(x_\sigma)$

Now, we have

$$FI(P_\sigma) = E_{x_\sigma \sim P_\sigma} [\|S_\sigma(x_\sigma)\|^2]$$

$$\begin{aligned} g(x) &= E[x_0 | x_\sigma = x] \\ &= \int_{-\infty}^{\infty} \omega \cdot p_{x_0 | x_\sigma}(w|x) dx \end{aligned}$$

$$= \frac{1}{\sigma^4} E_{x_\sigma \sim P_\sigma} \left[\left\| \underbrace{E[x_0 | x_\sigma]}_{g(x_\sigma)} - \underbrace{x_\sigma}_{x_0 + \sigma z} \right\|^2 \right]$$

$$\gamma = r(R, w)$$

$$E_\gamma[h(\gamma)]$$

$$= E_{R, w}[h(r(R, w))]$$

change of
variables

$$= \frac{1}{\sigma^4} E_{(x_0, z)} \left[\|g(x_0 + \sigma z) - (x_0 + \sigma z)\|^2 \right]$$

$$= \frac{1}{\sigma^4} E_{(x_0, z)} \left[\left\| \underbrace{(g(x_0 + \sigma z) - x_0)}_{e(x_0, z) \rightarrow \text{estimation error}} - \sigma z \right\|^2 \right]$$

$$= \frac{1}{\sigma^4} E_{(x_0, z)} \left[\|e(x_0, z)\|^2 + \sigma^2 \|z\|^2 - 2\sigma \langle e(x_0, z), z \rangle \right]$$

analyzed below

→ (a)

Now, consider: $E_{(x_0, z)} [\langle e(x_0, z), z \rangle]$

Again change variables: $(x_0, z) \rightarrow (x_0, x_\sigma)$

$$e(x_0, z) = (g(x_0 + \sigma z) - x_0) \\ \rightarrow \hat{e}(x_0, x_\sigma) = (g(x_\sigma) - x_0)$$

$$z \rightarrow \left(\frac{x_\sigma - x_0}{\sigma} \right)$$

Then: $E_{(x_0, z)} [\langle e(x_0, z), z \rangle]$

(change of variables)

$$= E_{(x_0, x_\sigma)} \left[\langle g(x_\sigma) - x_0, \frac{x_\sigma - x_0}{\sigma} \rangle \right]$$

$$= E_{(x_0, x_\sigma)} \left[\langle g(x_\sigma) - x_0, \underbrace{\frac{x_\sigma - g(x_\sigma)}{\sigma}}_{h(x_\sigma)} + \frac{g(x_\sigma) - x_0}{\sigma} \rangle \right]$$

Now, principle of orthogonality (see Theorem 3.2 in B. Hajek, Random Processes for Engineers, Cambridge Univ.

Press, 2014). implies that the error with the optimal mean-square estimator is orthogonal to any other estimate.

$$\text{i.e., } E_{(x_0, x_\sigma)} \left[\left\langle g(x_\sigma) - x_0, \frac{\overbrace{x_\sigma - g(x_\sigma)}^{h(x_\sigma)}}{\sigma} \right\rangle \right]$$

$$= E_{x_\sigma} \left[\underbrace{E_{x_0|x_\sigma} \left[\left\langle g(x_\sigma) - x_0, h(x_\sigma) \right\rangle \mid x_\sigma \right]}_{=0} \right] = 0$$

$$\therefore E_{(x_0, z)} \left[\left\langle e(x_0, z), z \right\rangle \right]$$

change of variables

$$= E_{(x_0, x_\sigma)} \left[\left\langle g(x_\sigma) - x_0, \frac{g(x_\sigma) - x_0}{\sigma} \right\rangle \right]$$

$$= \frac{1}{\sigma} E_{(x_0, x_\sigma)} \left[\|g(x_\sigma) - x_0\|^2 \right]$$

change of variables

$$= \frac{1}{\sigma} E_{(x_0, z)} \left[\|e(x_0, z)\|^2 \right]$$

→ (b)

Substituting (b) in (a), we have:

$$FI(P_\sigma) = \frac{1}{\sigma^4} E_{(x_0, z)} \left[\|e(x_0, z)\|^2 + \sigma^2 \|z\|^2 - 2 \langle e(x_0, z), z \rangle \right]$$

$$= \frac{1}{\sigma^4} E_{(x_0, z)} \left[-\|e(x_0, z)\|^2 + \sigma^2 \|z\|^2 \right]$$

$$= -\frac{1}{\sigma^4} \mathcal{E}(P_\sigma; \sigma) + \frac{d}{\sigma^2}$$

i.e.,

$$\mathcal{E}(P_\sigma; \sigma) = \sigma^2 d - \sigma^4 FI(P_\sigma).$$

This result is called Brown's identity [2, 3].

Another relation that is of interest is the **I-MMSE** [4].

We will use this later when studying time discretization with DDPM.

Mutual Information : Given r.v.s. (X, Y) with joint pdf $P_{X,Y}(x,y)$, and associated marginals $P_X(x)$ and $P_Y(y)$.

$$I(X; Y) = KL(P_{X,Y}(x,y) \parallel P_X(x) P_Y(y)).$$

Let $Y_\rho = \rho X + Z$, where $X \perp Z$
 $Z \sim N(0, 1)$.

Then,

$$\frac{d}{d\rho} I(X; Y_\rho) = \rho E[(X - E[X|Y_\rho])^2]$$

I-MMSE

MMSE error when
estimating X given Y_ρ .

Using methods similar to the previous derivation, we can derive the above; see [4] for discussion.

