$$\underline{f}^{MAP}(\underline{x}) = \underset{\underline{d}}{\operatorname{argmax}} \frac{P_{\underline{x}|\underline{D}}(\underline{x}|\underline{d})P_{\underline{D}}(\underline{d})}{P_{\underline{x}}(\underline{x})}$$

$$= \underset{\underline{d}}{\operatorname{argmax}} \underbrace{P_{\underline{x}|\underline{D}}(\underline{x}|\underline{d})}_{\text{likelihood}} \underbrace{P_{\underline{D}}(\underline{d})}_{\text{prior.}}$$

because $P_{\underline{x}}(\underline{x})$ does not depend on $\underline{d}$.

or Taking the log (an increasing function does not alter the order relationship)

$$\underline{f}^{MAP}(\underline{x}) = \underset{\underline{d}}{\operatorname{argmax}} \Big[\underbrace{\log P_{\underline{x}|\underline{D}}(\underline{x}|\underline{d})}_{\text{log-likelihood.}} + \log P_{\underline{D}}(\underline{d}) \Big]$$

The log transformation may be useful when the likelihood is in the exponential family of distributions (i.e. Gaussian).

ML ESTIMATION (MV MASSIMA VEROSIMIGLIANZA)

Note that the MAP estimator requires knowledge of both the prior $P_{\underline{D}}(\underline{d})$ and the likelihood $P_{\underline{x}|\underline{D}}(\underline{x}|\underline{d})$.

The prior $P_{\underline{D}}(\underline{d})$ represents our marginal knowledge of the hidden variable. When $P_{\underline{D}}(\underline{d})$ is not available, or its equivalently assumed to be uniform the estimator is

$$\underline{f}^{ML}(\underline{x}) = \underset{\underline{d}}{\operatorname{argmax}} P_{\underline{x}|\underline{D}}(\underline{x}|\underline{d})$$

MAXIMUM LIKELIHOOD ESTIMATOR

or taking the log

R.20

$$\underline{\hat{d}}^{ML}(\underline{x}) = \arg\max_{\underline{d}} \underbrace{\log p_{\underline{x}|\underline{D}}(\underline{x}|\underline{d})}_{\text{log-likelihood}}$$

In solving regression problems (cont. $\underline{D}$) ML is very popular because seldom we have knowledge of the prior on $\underline{D}$.

The argmax of $p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})$ could be found solving

$$\nabla_{\underline{d}}\, p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d}) = \underline{0}$$

or

likelihood equation

$$\nabla_{\underline{d}} \log p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d}) = \underline{0}$$

This condition must be verified to give a maximum value, that in general could be multiple.

Let us look now at some simple but typical examples of ML estimators

Example 1

$X = d + W$ ← Additive formation model

d: unknown constant

$$p_W(w) = \mathcal{N}(w; 0, \sigma_W^2)$$

$$p_{x|d}(x|d) = p_W(x-d) = \frac{1}{\sqrt{2\pi\sigma_W^2}} e^{-\frac{(x-d)^2}{2\sigma_W^2}}$$

$$\hat{d}^{ML} = \arg\max_{d} p_{x|d}(x|d) = x$$

Trivial result: take the observation as it is!!

EXAMPLE 2

$$\underline{x} = d\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} + \begin{bmatrix}w_1\\w_2\\\vdots\\w_N\end{bmatrix}$$

This is an "unknown" constant plus noise

Assume all w_i to be independent and distributed according to N densities $\{p_{w_i}(w_i), i=1,\dots,N\}$. Therefore the components of $\underline{x} = \begin{bmatrix}x_1\\x_2\\\vdots\\x_N\end{bmatrix}$ are independent, distributed according to the same densities of w_i but with mean d.

$$p_{\underline{x}|d}(\underline{x}|d) = \prod_{i=1}^{N} p_{w_i}(x_i - d)$$

Consider now various noise distributions.

2.1 GAUSSIAN NOISE $\quad p_{w_i}(w_i) = N(w_i, 0, \sigma_w^2)$

$$p_{\underline{x}|d}(\underline{x}|d) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{(x_i-d)^2}{2\sigma_w^2}}$$; Taking the log

$$\ln p_{\underline{x}|d}(\underline{x}|d) = \sum_{i=1}^{N}\left(\frac{1}{\sqrt{2\pi\sigma_w^2}} - \frac{(x_i-d)^2}{2\sigma_w^2}\right)$$

constant with respect to d

$$d^{ML} = \arg\max_d \ln p_{\underline{x}|d}(\underline{x}|d) = \arg\max_d -\sum_{i=1}^{N}\frac{(x_i-d)^2}{2\sigma_w^2}$$

$$= \arg\min_d \sum_{i=1}^{N}(x_i-d)^2$$

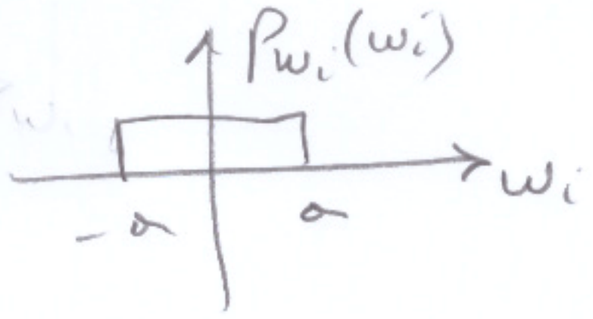
Taking the derivative w.r.t. d $\quad \sum_{i=1}^{N}(x_i-d)(-1) = 0$

$$\sum_{i=1}^{N} x_i - Nd = 0$$

$$\boxed{d^{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i}$$ Sample mean.

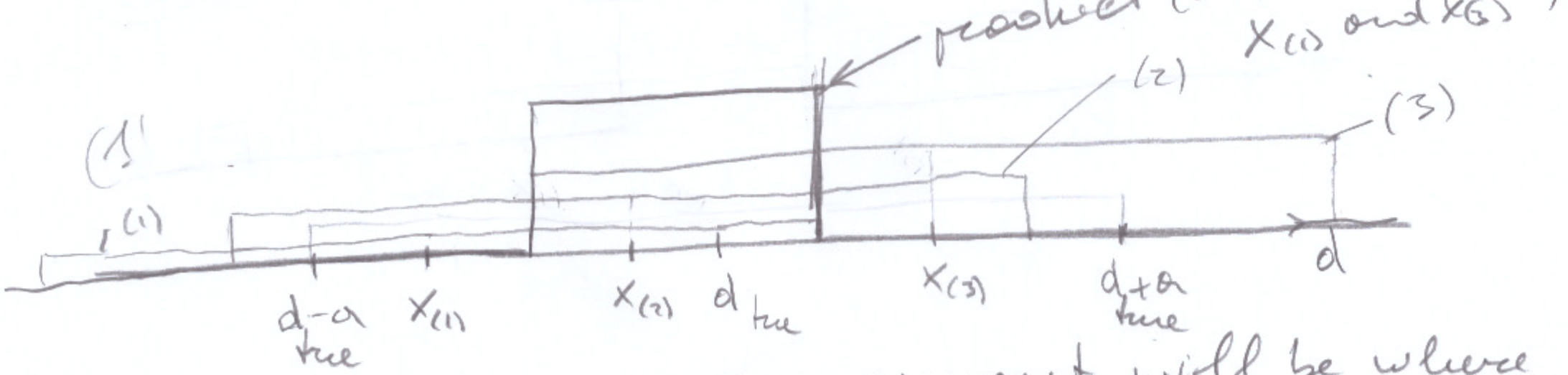
2.2 UNIFORM NOISE

$$p_{W_i}(w_i) = \mathcal{U}(-a, a) = \frac{1}{2a}\,\Pi\left(\frac{w_i}{2a}\right)$$



$$p_{\underline{x}|d}(\underline{x}|d) = \prod_{i=1}^{N} \Pi\left(\frac{x_i - d}{2a}\right)$$

Let us visualize the product will a picture for $N = 3$. Assume we get x_1, x_2 and x_3 that ranked are $x_{(1)} < x_{(2)} < x_{(3)}$



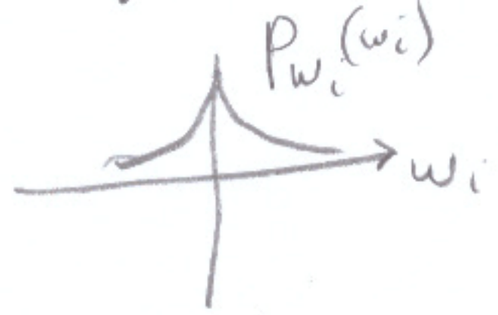
In general the non zero segment will be where all the functions are superimposed. But this is determined by the min $x_{(1)}$ and the max $x_{(N)}$

$$\boxed{\hat{d}^{ML} = \frac{x_{(1)} + x_{(N)}}{2}}$$

MIDRANGE OR MIN-MAX ESTIMATOR

2.3 LAPLACIAN NOISE (Double exponential)

$$p_{w_i}(w_i) = \frac{\lambda}{2} e^{-\lambda|w_i|}$$



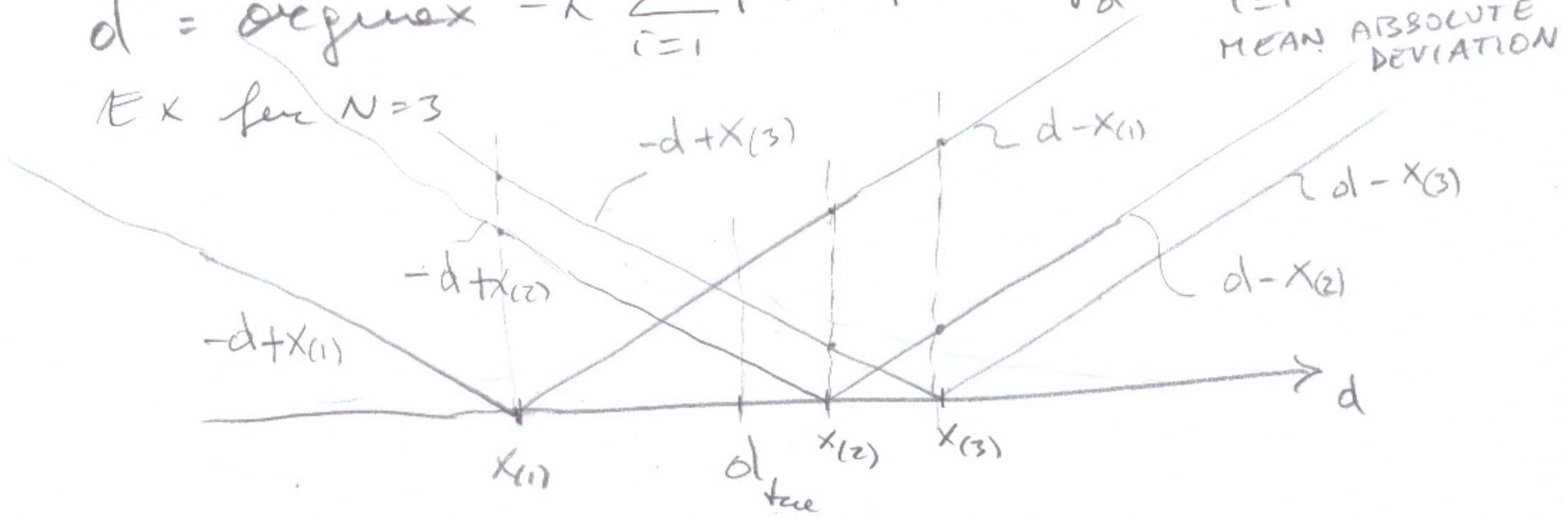
$$p_{\underline{x}|d}(\underline{x}|d) = \prod_{i=1}^{N} p_{w_i}(x_i - d) = \prod_{i=1}^{N} \frac{\lambda}{2} e^{-\lambda|x_i - d|}$$

$$\ln p_{\underline{x}|d}(\underline{x}|d) = \sum_{i=1}^{N} \left(\log\frac{\lambda}{2} - \lambda|x_i - d|\right)$$

$$d^{ML} = \arg\max -\lambda \sum_{i=1}^{N} |x_i - d| = \arg\min_d \sum_{i=1}^{N} |x_i - d|$$

MEAN ABSOLUTE DEVIATION

Ex for $N = 3$



Analyzing the four intervals

(a) $d < x_{(1)}$ $\quad -d + x_{(1)} - d + x_{(2)} - d + x_{(3)} = -3d + x_{(1)} + x_{(2)} + x_{(3)}$

(b) $x_{(1)} < d < x_{(2)}$ $\quad d - x_{(1)} - d + x_{(2)} - d + x_{(3)} = -d - x_{(1)} + x_{(2)} + x_{(3)}$

(c) $x_{(2)} < d < x_{(3)}$ $\quad d - x_{(1)} + d - x_{(2)} - d + x_{(3)} = d - x_{(1)} - x_{(2)} + x_{(3)}$

(d) $d > x_{(3)}$ $\quad d - x_{(1)} + d - x_{(2)} + d - x_{(3)} = 3d - x_{(1)} - x_{(2)} - x_{(3)}$

(a) $d_{minimum} = x_{(1)}$

(b) $d_{min} = x_{(2)}$

(c) $d_{min} = x_{(2)}$

(d) $d_{min} = x_{(3)}$

$d^{ML} = x_{(2)}$ Sample Median estimator

In general for N odd

$$d^{ML} = x_{\left(\frac{N+1}{2}\right)}$$

$x_{(1)}\ x_{(2)}\ x_{(3)}\ x_{(4)}$

for N even

$$d^{ML} = \frac{1}{2}\left(x_{\left(\frac{N}{2}\right)} + x_{\left(\frac{N}{2}+1\right)}\right)$$

EXAMPLE 3 NON LINEAR REGRESSION IN THE PRESENCE OF GAUSSIAN NOISE

$$\underline{x} = g(\underline{d}) + \underline{W}$$

$\underline{W} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix}$ Gaussian components zero mean and independent (same variance)

$p_{W_i}(w_i) = N(w_i, 0, \sigma_w^2)$

$$p_{\underline{x}|d}(\underline{x}|\underline{d}) = p_{\underline{W}}(\underline{x} - g(\underline{d}))$$

$$= \prod_{i=1}^{N} p_{W_i}(x_i - g_i(\underline{d})) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{(x_i - g_i(\underline{d}))^2}{2\sigma_w^2}}$$

$$\underline{d}^{ML} = \arg\max_{d} p_{\underline{x}|d}(\underline{x}|d) = \arg\max_{\underline{d}} \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})$$

$$= \arg\max_{\underline{d}} \sum_{i=1}^{N} \left(\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(x_i - g_i(\underline{d}))^2}{2\sigma_w^2}\right)$$

$$= \arg\min_{\underline{d}} \sum_{i=1}^{N} (x_i - g_i(\underline{d}))^2$$ sum of error squared

The solution could be found using the gradient

$$\nabla_{\underline{d}} \sum_{i=1}^{N} \left(x_i - g_i(\underline{d})\right)^2 = \underline{0}$$

$$-2 \sum_{i=1}^{N} \left(x_i - g_i(\underline{d})\right) \nabla_{\underline{d}} g_i(\underline{d}) = \underline{0}$$

When this equation cannot be solved, we could use Newton method using an iteration

$$\underline{d}(k) = \underline{d}(k-1) - \mu \nabla_{\underline{d}(k-1)}$$

$$d(k) = d(k-1) + \mu \sum_{i=1}^{N} \left(x_i - g_i(\underline{d}(k-1))\right) \nabla_{\underline{d}(k-1)} g_i(d(k-1))$$

given $x_1 x_2 \cdots x_N \longrightarrow$ Newton search d^{ML}

Example 4 GENERAL GAUSSIAN LIKELIHOOD

$$p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d}) = \mathcal{N}\left(\underline{x}; \underline{\mu}_{\underline{x}|\underline{d}}, \underline{\underline{\Sigma}}_{\underline{x}|\underline{d}}\right)$$

$$= \frac{1}{(2\pi)^{N/2}\left|\underline{\underline{\Sigma}}_{\underline{x}|\underline{d}}\right|^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_{\underline{x}|\underline{d}})^T \underline{\underline{\Sigma}}_{\underline{x}|\underline{d}}^{-1} (\underline{x}-\underline{\mu}_{\underline{x}|\underline{d}})}$$

$$\underline{d}^{ML} = \underset{\underline{d}}{\operatorname{argmax}}\; p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d}) = \underset{\underline{d}}{\operatorname{argmax}}\; \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})$$

$$= \underset{\underline{d}}{\operatorname{argmax}} \left[-\ln (2\pi)^{N/2} \left|\underline{\underline{\Sigma}}_{\underline{x}|\underline{d}}\right|^{1/2} - \frac{1}{2}(\underline{x}-\underline{\mu}_{\underline{x}|\underline{d}})^T \underline{\underline{\Sigma}}_{\underline{x}|\underline{d}}^{-1} (\underline{x}-\underline{\mu}_{\underline{x}|\underline{d}}) \right]$$

If the conditional covariance $\underline{\underline{\Sigma}}_{\underline{x}|\underline{d}}$ does not depend on $\underline{d}$ and is $\underline{\underline{\Sigma}}_x$

$$\underline{d}^{ML} = \underset{\underline{d}}{\operatorname{argmin}}\; (\underline{x}-\underline{\mu}_{\underline{x}|\underline{d}})^T \underline{\underline{\Sigma}}_x^{-1} (\underline{x}-\underline{\mu}_{\underline{x}|\underline{d}})$$

etc.

UNBIASED ESTIMATORS

An estimator of $\underline{D}$ is unbiased, if its mean matches the mean of the desired variable.

$$E[\underline{f}(\underline{x})] = E[\underline{D}]$$

When the variable is a constant, but unknown, $\underline{f}(\underline{x})$ is unbiased if

$$E[\underline{f}(\underline{x})] = \underline{d}$$

(the estimator's mean matches the true value)

MSE LOWER BOUND FOR UNBIASED ESTIMATORS

There exist a quite striking result that predicts a lower bound for the mean squared error estimators. This may be quite useful because, even if we do not have found the estimators, we may be able to predict a theoretical limit for the MSE. Consider first the scalar case

THEOREM (CRAMER-RAO LOWER BOUND):

If $f(x)$ is an unbiased estimate of d.

$$E[(f(x)-d)^2] \geq \frac{1}{E\left[\left(\frac{\partial \ln p_{x|d}(\underline{x}|d)}{\partial d}\right)^2\right]}$$

or equivalently

$$E[(f(x)-d)^2] \geq -\frac{1}{E\left[\frac{\partial^2 \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d^2}\right]}$$

with the following conditions satisfied

(c) $\frac{\partial p_{\underline{x}|d}(\underline{x}|d)}{\partial d}$ and $\frac{\partial^2 p_{\underline{x}|d}(\underline{x}|d)}{\partial d^2}$

exist and are absolutely integrable.

Proof: Since $f(\underline{X})$ is unbiased, we have

$$E[f(\underline{X}) - d] \triangleq \int_X (f(\underline{x}) - d)\, p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} = 0$$

Differentiating with respect to d, we have

$$\frac{\partial}{\partial d} \int_X (f(\underline{x}) - d)\, p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} = 0$$

Integrability is ensured by the hypothesis of the theorem, therefore we can write

$$\int_X \frac{\partial}{\partial d}\left[(f(\underline{x}) - d)\, p_{\underline{x}|d}(\underline{x}|d)\right] d\underline{x} = 0$$

$$-\int_X p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} + \int_X \left[(f(\underline{x}) - d)\, \frac{\partial}{\partial d} p_{\underline{x}|d}(\underline{x}|d)\right] d\underline{x} = 0$$

($\int_X p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} = 1$)

Now observe that

$$\frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d} = \frac{1}{p_{\underline{x}|d}(\underline{x}|d)} \cdot \frac{\partial p_{\underline{x}|d}(\underline{x}|d)}{\partial d}.$$

Extracting $\frac{\partial p_{\underline{x}|d}}{\partial d}$ and substituting it into the above equation we get:

CRLB.3

$$\int_{\underline{x}} \frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\, p_{\underline{x}|d}(\underline{x}|d)\,(f(\underline{x})-d)\, d\underline{x} = 1$$

We can rewrite it as

$$\int_{\underline{x}} \left[\frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\sqrt{p_{\underline{x}|d}(\underline{x}|d)}\right]\left[\sqrt{p_{\underline{x}|d}(\underline{x}|d)}\,(f(\underline{x})-d)\right] d\underline{x} = 1$$

Using Schwartz's inequality

$$\left|\int f(\underline{x})\, g(\underline{x})\, d\underline{x}\right|^2 \le \int f^2(\underline{x})\, d\underline{x} \cdot \int g^2(\underline{x})\, d\underline{x}$$

$$1 \le \int_{\underline{x}} \left[\frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\right]^2 p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} \cdot \int_{\underline{x}} (f(\underline{x})-d)^2\, p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x}$$

Therefore

$$E[(f(x)-d)^2] \ge \frac{1}{E\left[\left(\frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\right)^2\right]}$$

the equality in Schwartz inequality holds iff $f(x) = K g(x)$ for some K. In our case the equality holds iff.

$$K(f(\underline{x})-d) = \frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d} \quad \text{for some } K.$$

To prove form (6), observe that

$$\int_{\underline{x}} p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} = 1$$

Differentiating with respect to d, we have

$$\int_{\underline{x}} \frac{\partial\, p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\, d\underline{x} = 0$$

or

$$\int_{\underline{x}} \frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\, p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} = 0$$

Differentiating again with respect to d, we have

$$\int_{\underline{x}} \frac{\partial^2 \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d^2}\, p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x} + \int \left(\frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\right)^2 p_{\underline{x}|d}(\underline{x}|d)\, d\underline{x}$$

$$= 0\,,$$

which can be written as

$$E\left[\left(\frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d}\right)^2\right] = -E\left[\frac{\partial^2 \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d^2}\right].$$

the above expression immediately gives us (b) ▨

Note that only if the condition

$$K(d)\big(f(\underline{x}) - d\big) = \frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d} \qquad (*)$$

is satisfied, the lower bound is achieved. Note that the constant K can depend on d and that in general this may be a rather restrictive condition. Nevertheless if the condition holds and $\frac{\partial \ln p_{\underline{x}|d}(\underline{x}|d)}{\partial d} = 0$,

which corresponds to a unique solution for the CRLB.5
ML estimator, the ML estimator reaches the lower bound. It is rare for general cases that estimator reach the lower bound.

EXAMPLE case 1

Take example 2 with W gaussian. Namely

$$\underline{X} = d\begin{bmatrix}1\\1\\1\end{bmatrix} + \underline{W} \quad .$$

the ML equation reads

$$\frac{\partial}{\partial d} \ln p_{\underline{x}|d}(\underline{x}|d) = \frac{1}{\sigma_w^2}\left(\sum_{i=1}^{N} x_i - d\right) = 0$$

or

$$\frac{\partial}{\partial d} \ln p_{\underline{x}|d}(\underline{x}|d) = \frac{N}{\sigma_w^2}\left(\frac{1}{N}\sum_{i=1}^{N} x_i - d\right) = 0$$

the first part of the equation satisfies the regularity condition (*), therefore the bound is achieved. the ML estimator, (sample mean)

$$\hat{d}_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i \ ,$$

gives

$$E\left[(f(\underline{x}) - d)^2\right] = \frac{\sigma_w^2}{N}$$

since

$$\frac{\partial^2}{\partial d^2} \ln p_{\underline{x}|d}(\underline{x}|d) = -\frac{N}{\sigma_w^2} \quad .$$

and because it is unbiased:

$$E\left[\frac{1}{N}\sum_{i=1}^{N} x_i\right] = \frac{1}{N}\sum_{i=1}^{N} E[d + W_i] = d \quad .$$

THE MULTIDIMENSIONAL CASE

THEOREM Consider any unbiased estimate $\underline{f}(\underline{x})$ of $\underline{d}$.

Then $$E\left[(f_i(\underline{x}) - d_i)^2\right] \geq J^{ii} \qquad \forall i = 1, \dots, M$$

where J^{ii} the ith diagonal element of the $L \times L$ square matrix $\underline{\underline{J}} = \underline{\underline{F}}^{-1}$

where

$$[F]_{ij} = \left[E\left[\frac{\partial \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_i} \cdot \frac{\partial \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_j}\right]\right]_{i,j=1,\dots,M}$$

which by the property proved above (*) can be written as

$$[F]_{ij} = \left[-E\left[\frac{\partial^2 \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_i \partial d_j}\right]\right]_{i,j}.$$

The matrix $\underline{\underline{F}}$ is called FISHER INFORMATION MATRIX. In fact it measures the sensitivity of our observations (in probability) with respect to variation in the observed estimated quantity.

In matrix notation, the Fisher's information matrix can also be expressed as

---

(*) To prove this equivalence we can just differentiate twice, once with respect to d_i and once with respect to d_j the equation $\int_{\underline{x}} p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})\, d\underline{x} = 1$, and by noting that $\frac{\partial \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_i} = \frac{1}{p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})} \frac{\partial p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_i}$

$$\underline{\underline{F}} = E\left[\left(\nabla_{\underline{d}} \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})\right)\left(\nabla_{\underline{d}} \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})\right)^T\right]$$

$$= -E\left[\left(\nabla_{\underline{d}} \nabla_{\underline{d}}^T\right) \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})\right]$$

The bound equality holds only if

$$\left(f_i(\underline{x}) - d_i\right) = \sum_{j=1}^{L} k_{ij}(\underline{d}) \frac{\partial \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_j},$$

for all $i = 1, \ldots, M$, and $\forall \underline{x} \in \mathcal{X}$.

PROOF. The proof follows the same guidelines of that for the one-dimensional case: Since the estimator is assumed to be unbiased, we have

$$\int_{\mathcal{X}} \left(f_i(\underline{x}) - d_i\right) p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})\, d\underline{x} = 0 \quad , \forall i = 1, \ldots, M,$$

or

$$\int_{\mathcal{X}} f_i(\underline{x})\, p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})\, d\underline{x} = d_i \quad , \forall i = 1, \ldots, M$$

Differentiating both sides with respect to d_j

$$\int_{\mathcal{X}} f_i(\underline{x}) \frac{\partial p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_j} d\underline{x} = \int_{\mathcal{X}} f_i(\underline{x}) \frac{\partial \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_j} p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})\, d\underline{x}$$

$$= \delta_{ij} \quad , \forall j = 1, \ldots, M.$$

δ_{ij} is the Kronecker delta: $\delta_{ij} \triangleq \begin{cases} 1 & i = j \\ 0 & \text{else.} \end{cases}$

CRLB.8

let us prove the result for $i=1$. The same procedure would follow for any $i=1,\dots,M$

let us define the $(M+1)$-dimensional vector

$$\underline{X} = \begin{bmatrix} f_1(\underline{x}) - d_1 \\ \dfrac{\partial \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_1} \\ \vdots \\ \dfrac{\partial \ln p_{\underline{x}|\underline{d}}(\underline{x}|\underline{d})}{\partial d_L} \end{bmatrix}$$

the covariance matrix of $\underline{\alpha}$ is:

$$E[\underline{\alpha}\underline{\alpha}^T] = \left[\begin{array}{c|ccc} \sigma_{\varepsilon_1}^2 & 1 & 0 \dots & 0 \\ \hline 1 & & & \\ \vdots & & \underline{\underline{F}} & \\ 0 & & & \end{array}\right].$$

the ones come from the above expression, and $\sigma_{\varepsilon_1}^2$ is just the mean square error on the first variable. the matrix is non negative definite since it is a covariance matrix. Evaluating the determinant using the cofactor expansion we get

$$\sigma_{\varepsilon_1}^2 |\underline{\underline{F}}| - \text{cofactor } F_{11} \geq 0 .$$

therefore if $\underline{\underline{F}}$ is non singular

$$\sigma_{\varepsilon_1}^2 \geq \frac{\text{cofactor } F_{11}}{|F|} = J_{11} \qquad \text{CRLB.9}$$

where J_{11} is the first element of $\underline{\underline{F}}^{-1}$.

The case when $\underline{\underline{F}}$ is singular has to be worked out for every specific problem.

In order for the determinant to be equal to zero, the term $f_1(\underline{x}) - d_1$ must be expressible as a linear combination of the other terms in $\underline{x}$. This is the condition for equality.

Exam

RANDOM PARAMETER $\underline{D}$

Clearly if some a-priori information is given about $\underline{D}$, it should be included in the bound. The condition on an estimator to be unbiased if $\underline{D}$ is a random L-dimensional variable is

$$E[\underline{\hat{f}}(\underline{X})] = E[\underline{D}]$$

or

$$E[\underline{\hat{f}}(\underline{X}) - \underline{D}] = \int_{X \times D} (\underline{\hat{f}}(\underline{x}) - \underline{d}) \, p_{\underline{X}\underline{D}}(\underline{x}, \underline{d}) \, d\underline{x} \, d\underline{d}$$

or

$$\int_{X \times D} \underline{f}(\underline{x}) \, p_{\underline{X}|\underline{D}}(\underline{x}|\underline{d}) \, p_{\underline{D}}(\underline{d}) \, d\underline{x} \, d\underline{d} = \int_{D} \underline{d} \, p_{\underline{D}}(\underline{d}) \, d\underline{d} \; .$$

Rewriting the equation component-by component

$$\int_{X \times D} f_i(\underline{x}) \, p_{\underline{X}|\underline{D}}(\underline{x}|\underline{d}) \, p_{\underline{D}}(\underline{d}) \, d\underline{x} \, d\underline{d} = \int_{D} d_i \, p_{\underline{D}}(\underline{d}) \, d\underline{d}$$

$\forall i = 1, \ldots, L$.

Differentiating both sides with respect to d_j, with the usual assumptions that allow to take the derivatives inside the integration signs, we have

$$\int_{X \times D} f_i(\underline{x}) \left[\frac{\partial \, p_{\underline{X}|\underline{D}}(\underline{x}|\underline{d})}{\partial d_j} \cdot p_{\underline{D}}(\underline{d}) + p_{\underline{X}|\underline{D}}(\underline{x}|\underline{d}) \frac{\partial \, p_{\underline{D}}(\underline{d})}{\partial d_j} \right] d\underline{x} \, d\underline{d} =$$

$$= \int_D \left(\delta_{ij} \, p_{\underline{D}}(\underline{d}) + \underline{d} \, \frac{\partial p_{\underline{D}}(\underline{d})}{\partial d_j} \right) d\underline{d}$$

. . . .

carrying out all the derivations, in a way similar to the procedure used for the case of $\underline{D}$ non-random we get that the same result holds, but the Fisher's information matrix has to be re-defined as

$$F = F_1 + F_2 \,,$$

F_1 has the usual definition given above, and

$$\left[F_2 \right]_{ij} = E\left[\frac{\partial \ln p_{\underline{D}}(\underline{d})}{\partial d_i} \; \frac{\partial \ln p_{\underline{D}}(\underline{d})}{\partial d_j} \right]$$

or

$$\left[F_2 \right]_{ij} = - E\left[\frac{\partial^2 \ln p_{\underline{D}}(\underline{d})}{\partial d_i \, \partial d_j} \right]$$

$\forall \, i, j = 1, \ldots, M$.

(see Van Trees for more details)