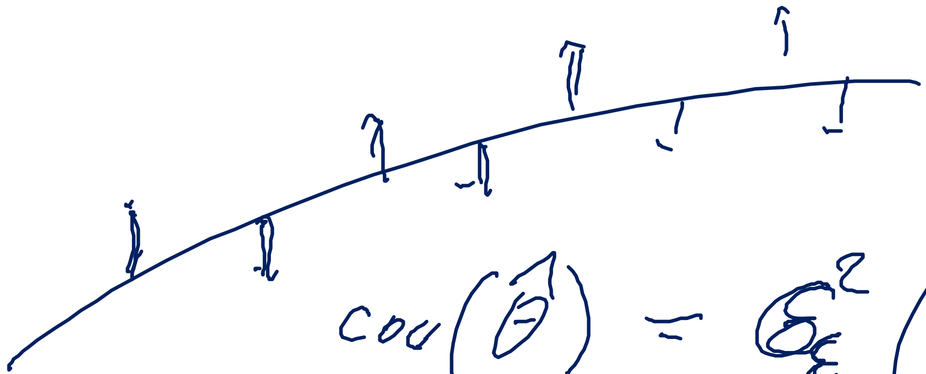


$$\sigma_\varepsilon^2 = \frac{1}{N-d}\sum_{k=1}^{N}\left(y(k)-\phi(k)^T\underline{\underline{\hat{\theta}}}\right)^2$$

$$\mathrm{cov}(\hat{\theta}) = \sigma_\varepsilon^2\,\underline{\underline{\left(\phi_N^T\phi_N\right)^{-1}}}$$

CH 5: MAXIMUM LIKELIHOOD AND BAYESIAN ESTIMATION

MODEL:

$M: \mathbb{R}^d \to \mathbb{R}^N$

$y, \varepsilon \in \mathbb{R}^N$

$$\boxed{y = M(\theta) + \varepsilon}$$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$$

$M(\theta)$ DETERMINISTIC, ε ZERO MEAN, . . .
(INDEPENDENT)
(I.I.D.)

5.1 MAX. LIKELIHOOD (ML)

DEF: $L(\theta) = p(y|\theta)$ LIKELIHOOD FUNCTION

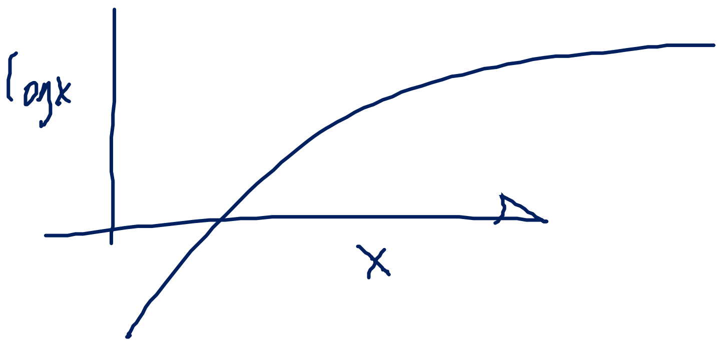
"PDF OF y GIVEN θ FOR y FIXED"

DEF: ML-ESTIMATOR: GIVEN Y, MAXIMIZE $p(Y|\theta)$ WRT. θ

EX: ε_i, $\varepsilon_1, \dots, \varepsilon_N$. ZERO MEAN, GAUSSIAN(σ_i^2)
INDEPENDENT

$$p(y|\theta) = \prod_{i=1}^{N} p(y_i|\theta) = C \prod_{i=1}^{N} \exp\left(-\frac{(y_i - M_i(\theta))^2}{2\sigma_i^2}\right)$$

$$-\log p(y|\theta) = -\log C + \sum_{i=1}^{N} \frac{(y_i - M_i(\theta))^2}{2\sigma_i^2}$$



$$\boxed{\arg\max\, p(y|\theta) = \arg\min\, -\log p(y|\theta)}$$

MAX. LIKELIHOOD $\hat{=}$ NONLINEAR LEAST SQUARES

$$\sum_{i=1}^{N} \frac{(y_i - M_i(\theta))^2}{2\sigma_i^2} = \frac{1}{2}\left\| S^{-1}\left(y - M(\theta)\right)\right\|_2^2$$

$$= \frac{1}{2}\left\| \underbrace{S^{-1}y - S^{-1}\cdot M(\theta)}\right\|_2^2$$

$$= \frac{1}{2}\left\| F(\theta) \right\|_2^2$$

$$S = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_N \end{pmatrix}$$

$$S = \Sigma^{\frac{1}{2}}$$

→ SOLVE WITH "lsqnonlin" IN MATLAB (FINDS LOCAL MINIMUM) (GAUSS-NEWTON METHOD)

MORE GENERAL, CAN TREAT CORRELATED NOISE $\varepsilon \sim \mathcal{N}(0, \Sigma)$

$$p(y|\theta) = C \cdot \exp\left(- \frac{(y - M(\theta))^T \Sigma^{-1} (y - M(\theta))}{2}\right)$$

$$-\log p(y|\theta) = \text{CONST} + \underbrace{\frac{1}{2}(y - M(\theta))^T \Sigma^{-1}(y - M(\theta))}_{= \frac{1}{2}\left\| \Sigma^{-\frac{1}{2}} y - \Sigma^{-\frac{1}{2}} M(\theta)\right\|_2^2}$$

REGULARIZATION: ADDITION OF AD-HOC / A-PRIORI KNOWLEDGE CAN HELP TO REDUCE NON-UNIQUENESS

$$\min \frac{1}{2}\|y-M(\theta)\|_2^2 + \frac{\alpha}{2}\|\theta-\bar{\theta}\|_2^2$$

WITH α SMALL

"PSEUDO-MEASUREMENT"

($\rightarrow$ BAYESIAN ESTIMATION)

5.1.1 L1-ESTIMATION

NON-SMOOTH

$$\boxed{\sum_{i=1}^{N} \frac{|y_i - M_i(\theta)|}{2a_i}}$$

"ROBUST ESTIMATION"

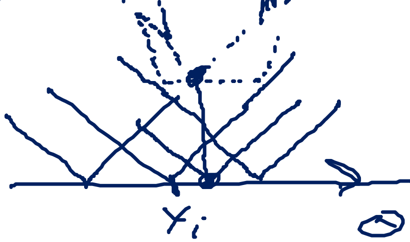
(LAPLACE)

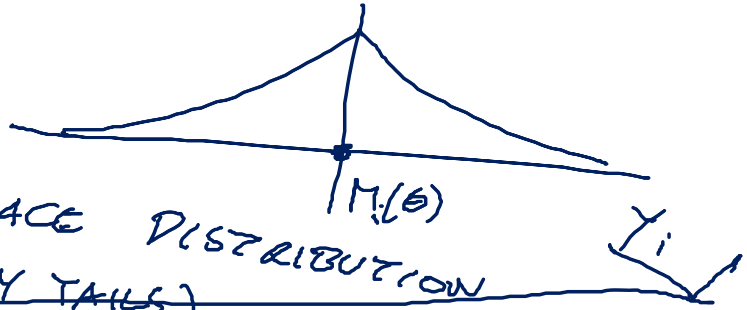
$$p(y_i|\theta) = C \cdot \exp\left(-\frac{|y_i - M_i(\theta)|}{2a_i}\right)$$

EXAMPLE: L1 $M_i(\theta) = \theta \in \mathbb{R}$, $a_i \equiv a$

$$\boxed{\arg\min_{\theta} \sum |y_i - \theta| = \text{MEDIAN}}$$

(ROBUST AGAINST OUTLIERS)

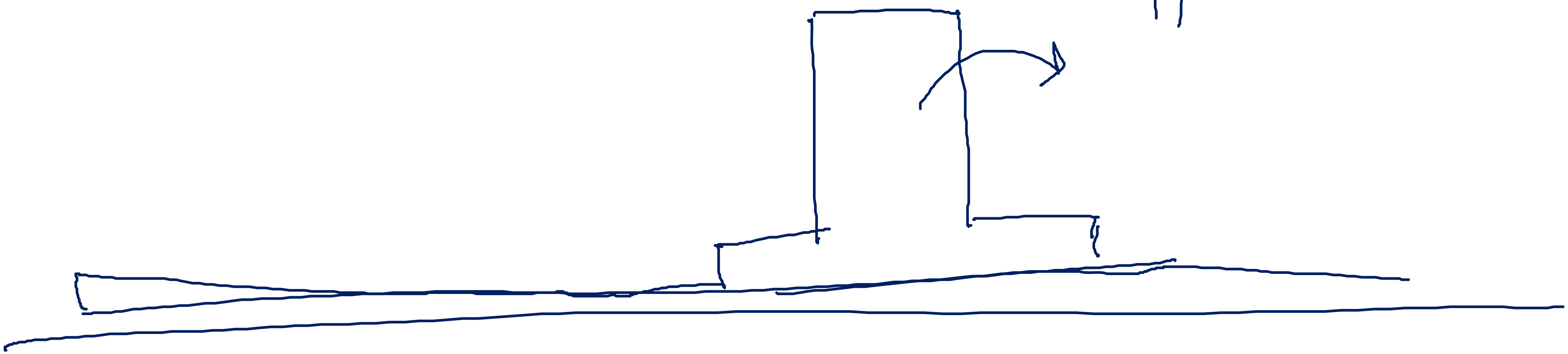




LAPLACE DISTRIBUTION (HEAVY TAILS)

READING SCRIPT

NOT STARTED YET	CH. 1 (INTRO READ, BUT NOT CH2	CH 2 (PROB)	CH3 (OPT)	CH4 (CLLS)	MORE
1	1	4	21	4	0



WORKLOAD PER WEEK (IN ADDITION TO PRESENCE IN LECT. & EX.)
(AVERAGE UNTIL NOW)

1h	2h	4h	6h	8h	10 OR MORE
0	0	2	6	13	15

~15

WHAT IS TIME USED FOR (50% OR MORE)

EXERCISE SHEETS	READING SCRIPT
MAJORITY	0

MATLAB	PAPER
13	~20

OTHER SOURCES

PROB. & STATISTICS TEXTBOOK

5.2 BAYESIAN ESTIMATION

INSTEAD OF $p(y|\theta)$ WE MAY WANT TO MAXIMIZE $p(\theta|y)$

"θ GIVEN y"

"BAYES FORMULA"

ML-TERM

A-PRIORI

$$\boxed{p(\theta|y) \left[= \frac{p(\theta, y)}{p(y)}\right] = \frac{p(y|\theta)\cdot p(\theta)}{p(y)}}$$

CONST. W.R.T. θ

DEF:

"MAXIMIZE $p(\theta|y)$" $\triangleq$ "MAXIMUM A-POSTERIORI (MAP) ESTIMATOR"

$$\arg\max_{\theta} p(\theta|y) = \arg\min_{\theta}\left(-\log p(\theta|y)\right) = \arg\min_{\theta}\Big(\underbrace{-\log p(y|\theta)}_{\text{ML-ESTIMATION}} \underbrace{-\log p(\theta)}_{\text{"REGULARIZATION" A-PRIORI INFO.}}\Big)$$

NOTE: A-PRIORI INFORMATION LEADS TO BIAS

RELATION BETWEEN ML & MAP

ML IS SPECIAL CASE OF MAP, BUT WITH ZERO A-PRIORI INFORMATION

MAP CAN BE CONS. SPECIAL CASE OF ML, BUT WITH "PSEUDO-MEASUREMENT"