ESTIMAREA CU PLAUZIBILITATE MAXIMĂ, ML

În unele situații fie nu putem să găsim un estimator MVU, fie el nu există. În astfel de situații putem să recurgem la estimarea cu plauzibilitate (verosimilitate) maximă. Estimatorii ce rezultă dintr-o astfel de abordare sunt cei mai răspândiți în aplicațiile practice, în rezolvarea unor probleme complicate de estimare. În multe cazuri, performanțele acestor estimatori sunt aproape optimale, dacă înregistrările de date sunt foarte lungi. În consecință, estimatorii ce maximizează plauzibilitatea, sunt aproximări bune ale estimatorilor MVU. Dacă estimatorul este suboptimal dar

$$\lim_{N\to\infty} E\left\{\hat{\theta}\right\} = \theta$$

$$\lim_{N\to\infty} Disp\left\{\hat{\theta}\right\} = CRLB_\theta = I^{-1}(\theta)$$

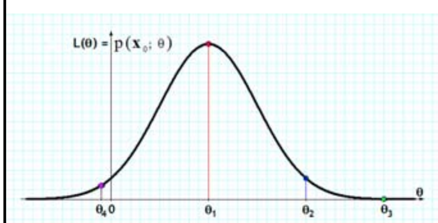
se spune că estimatorul este asimptotic nedeplasat, şi asimptotic eficient. Pentru înregistrări lungi, adică *N* destul de mare, un astfel de estimator poate fi considerat ca fiind optimal.

1

Determinarea estimatorului de plauzibilitate maximă

Estimatorul de plauzibilitate maximă, ML, este acea valoare a parametrului necunoscut, θ, care maximizează p(**x**; θ), cu **x** fixat, ceeace înseamnă că maximizează funcția de plauzibilitate, L(θ)

$$\hat{\theta}_{ML} = \arg\max_{\theta} L(\theta) = \max_{\theta} p\left(\mathbf{x} = \mathbf{x_0}; \theta\right)$$



Probabilitatea ca **x** observat, să cadă într-un volum Δ**x** ce tinde spre 0, centrat pe $\mathbf{x}_0$ este p(**x**=$\mathbf{x}_0$; θ)Δ**x** şi ea depinde, în principiu, de valoarea adevărată a parametrului θ. În figură se arată p(**x**=$\mathbf{x}_0$; θ) ca funcție de θ. Dacă valoarea adevărată a parametrului θ ar fi θ_3 atunci probabilitatea de a se fi observat **x**=$\mathbf{x}_0$ ar fi foarte mică, practic nulă. Prin urmare, ipoteza $\theta=\theta_3$ este una puțin credibilă, puțin plauzibilă. Ipotezele $\theta=\theta_2$ sau $\theta=\theta_4$, în condițiile în care s-a observat că **x**=$\mathbf{x}_0$, vor fi respinse pe aceleași motive. Urmând acelaşi raționament, ipoteza $\theta=\theta_1$ este cea mai plauzibilă, deoarece p(**x**=$\mathbf{x}_0$; $\theta=\theta_1$) are valoarea maximă.

Pentru exemplificare vom considera un semnal cu modelul

$$x[n] = A + w[n]; \quad A > 0,\ w[n] \sim \mathcal{N}(0,\ A) \quad n = 0,1,\ldots,N-1 \quad x[n] \sim \mathcal{N}(A,\ A)$$

A fiind atât valoarea componentei continue cât și dispersia zgomotului alb, gaussian

2

Densitatea de probabilitate a datelor are expresia

$$p(\mathbf{x};A)=\frac{1}{(2\pi A)^{N/2}}\exp\left\{-\frac{1}{2A}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\};\ \ A>0$$

Dacă datele sunt fixate, adică cele N componente ale vectorului **x** se înlocuiesc cu cele N valori efectiv măsurate, se obține funcția de plauzibilitate, o funcție ce depinde doar de parametrul necunoscut, A. În formă logaritmică funcția este

$$l(A)=\ln p(\mathbf{x};A)=-\frac{N}{2}\ln(2\pi)-\frac{N}{2}\ln A-\frac{1}{2A}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2;\mathbf{x}\text{ fixat, } A>0$$

Maximul ei apare pentru A>0, pentru care derivata este nulă

$$\frac{dl(A)}{dA}=-\frac{N}{2A}+\frac{1}{2A^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2+\frac{1}{A}\sum_{n=0}^{N-1}\left(x[n]-A\right)=0$$

Dezvoltând pătratul și efectuînd calculele, se obține ecuația

$$-\frac{N}{2A}+\frac{1}{2A^2}\sum_{n=0}^{N-1}x[n]^2-\frac{1}{A}\sum_{n=0}^{N-1}x[n]+\frac{N}{2}+\frac{1}{A}\sum_{n=0}^{N-1}x[n]-N=0$$

3

sau

$$A^2+A-\frac{1}{N}\sum_{n=0}^{N-1}x[n]^2=0;\qquad A>0$$

Soluțiile ecuației sunt

$$A_{1,2}=-\frac{1}{2}\pm\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x[n]^2+\frac{1}{4}}$$

dar numai soluția pozitivă are sens în exemplu

$$\hat{A}_{ML}=-\frac{1}{2}+\sqrt{\frac{1}{N}\sum_{n=0}^{N-1}x[n]^2+\frac{1}{4}}$$

4

Vom considera modelul de semnal în care nu cunoaştem componenta continuă A, zgomotul w[n] fiind alb, gaussian cu dispersia cunoscută

$$x[n] = A + w[n]; \quad n = 0,1,\dots,N-1; \quad w[n] \sim \mathcal{N}\left(0, \ \sigma^2\right)$$

Densitatea de repartiție a datelor **x** este

$$p(\mathbf{x}; A) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}$$

Cu **x** fixat (adică înlocuit cu vectorul obținut prin măsurare) plauzibilitatea logaritmică, dependentă doar de parametrul necunoscut, A, este

$$l(A) = \ln p(\mathbf{x}; A) = -\frac{N}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2; \quad \mathbf{x} \ \text{ fixat}$$

Ecuația ce permite daterminarea estimatului ML pentru A se obține anulând derivata, în raport cu A, a plauzibilității logaritmice

$$\frac{\partial l(A)}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1} x[n] - NA = 0$$

Se obține estimatul ML, care în acest caz, este chiar media eșantion

$$\hat{A}_{ML} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] = \overline{x}$$

Dacă într-o problemă există un estimator eficient, atunci procedura ML îl va produce !

5

Proprietăți asimptotice ale estimatorilor ML

Estimatorii ML sunt, asimptotic, nedeplasați și eficienți, adică

$$\lim_{N\to\infty} E\left\{\hat{\theta}\right\} = \theta; \quad \lim_{N\to\infty} Disp\left\{\hat{\theta}\right\} = CRLB_\theta = I^{-1}(\theta)$$

Simbolic, distribuția asimptotică a estimatorului ML se notează cu

$$\hat{\theta} \overset{a}{\sim} \mathcal{N}\left(\theta, \ I^{-1}(\theta)\right)$$

Pentru primul exemplu, în care atât media cât și dispersia datelor este A, primele două derivate ale plauzibilității logaritmice sunt

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A}\sum_{n=0}^{N-1}\left(x[n]-A\right) + \frac{1}{2A^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2$$

$$\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A^2} = \frac{N}{2A^2} - \frac{1}{A^2}\sum_{n=0}^{N-1} x[n] - \frac{1}{A^2}\sum_{n=0}^{N-1}\left(x[n]-A\right) - \frac{1}{A^3}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2$$

Mediind statistic a doua derivată obținem

$$E\left\{\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A^2}\right\} = \frac{N}{2A^2} - \frac{1}{A^2}\sum_{n=0}^{N-1} E\left\{x[n]\right\}$$

$$-\frac{1}{A^2}\sum_{n=0}^{N-1}\left(E\left\{x[n]\right\}-A\right) - \frac{1}{A^3}\sum_{n=0}^{N-1} E\left\{\left(x[n]-A\right)^2\right\}$$

$$= \frac{N}{2A^2} - \frac{N}{A} - \frac{N}{A^2} = -\left(\frac{N}{2A^2} + \frac{N}{A}\right)$$

6

În consecință inversa informației Fisher este

$$I^{-1}(A) = -\frac{1}{E\left\{\frac{\partial^2 \ln p(\mathbf{x};A)}{\partial A^2}\right\}} = \frac{1}{\frac{N}{2A^2}+\frac{N}{A}} = \frac{1}{N}\frac{A^2}{A+\frac{1}{2}}$$

Am stabilit deci repartiția asimptotică a estimatorului ML pentru A

$$\hat{A}_{ML} \overset{a}{\sim} \mathcal{N}\left(A,\ \frac{1}{N}\frac{A^2}{A+1/2}\right)$$

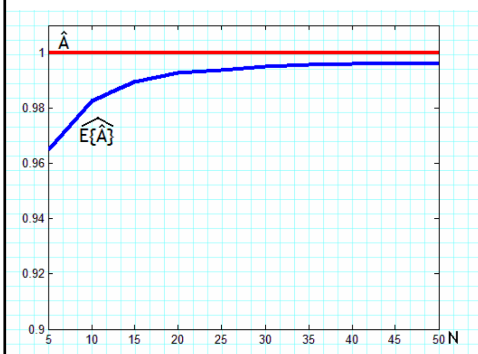
Pentru A=1 avem,

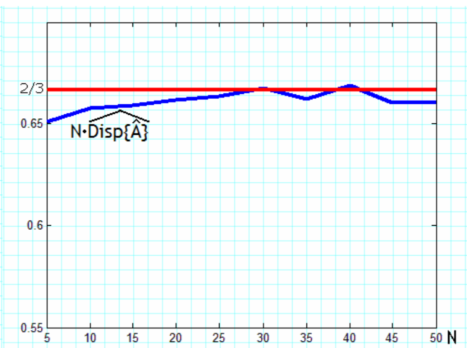
$$NI^{-1}(A)\Big|_{A=1} = \frac{1}{1+1/2} = \frac{2}{3}$$

produs constant, independent de N. Pentru a ne forma o idee despre ce înseamnă “N suficient de mare” am realizat un experiment pentru lungimi tot mai mari ale datelor, N=5, 10, 15, ...,50. Am repetat de M=10000 de ori experimentul, calculând 10000 de valori ale mediei și ale dispersiei. Pentru a reduce fluctuațiile statistice am calculat medii ale mediilor și ale dispersiilor obținând estimări ale medie statistice și ale dispersiei (statistice)

$$\hat{E}\{\hat{A}\} = \frac{1}{M}\sum_{i=1}^{M}\hat{A}_i;\quad \overset{\wedge}{Disp}\{\hat{A}\} = \frac{1}{M}\sum_{i=1}^{M}\left(\hat{A}_i - \hat{E}\{\hat{A}\}\right)^2;\quad M = 10.000$$

7





Se observă că estimarea mediei ne sugerează că pentru N>25 ne-am apropiat suficient de valoarea adevărată A=1 pentru a declara că estimatorul ML este nedeplasat

Produsul N.Dispersia este practic la valoarea 2/3 pentru N>25

În mod uzual, N suficient de mare pentru a admite că este adevărată repartiția asimptotică, înseamnă câteva zeci de eșantioane

8

Putem da, fără demonstrație, teorema

Dacă densitatea de probabilitate a datelor **x**, p(**x**; θ), satisface condiția de regularitate

$$E\left\{\frac{\partial p(\mathbf{x};\theta)}{\partial\theta}\right\}=0$$

atunci estimatorul pentru parametrul necunoscut θ, ce maximizează plauzibilitatea, adică estimatorul ML, este distribuit asimptotic conform cu

$$\hat{\theta}\overset{a}{\sim}\mathcal{N}\left(\theta,\quad I^{-1}(\theta)\right)$$

I(θ) fiind informația Fisher, evaluată la valoarea adevărată a parametrului θ.

***

Se poate deci spune că, în mod asimptotic, estimatorul ML este nedeplasat şi eficient, adică el este asimptotic optimal. Viteza de convergență, adică valoarea *N* după care se pot aplica formulele asimptotice, este o problemă deschisă. În majoritatea cazurilor *N* nu este foarte mare, fiind, în mod uzual, de ordinul zecilor.

9

Estimator ML pentru faza inițială a unei sinusoide

Reluăm problema determinării fazei inițiale pentru o sinusoidă, afectată de un zgomot alb, gaussian. Modelul datelor este

$$x[n]=A\cos(2\pi f_0 n+\Phi)+w[n];\quad w[n]\sim\mathcal{N}\left(0,\quad\sigma^2\right);\quad n=0,1,\ldots,N-1$$

Dacă se fixează **x** în expresia densității de probabilitate a datelor, se obține plauzibilitatea, funcție numai de parametrul necunoscut, Φ

$$L(\Phi)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left[x[n]-A\cos(2\pi f_0 n+\Phi)\right]^2\right\};\quad \mathbf{x}\text{ fixat}$$

Pentru stabilirea estimatorului ML e necesar să maximizăm L(Φ), ceeace e echivalent cu a minimiza, după Φ, funcția

$$J(\Phi)=\sum_{n=0}^{N-1}\left[x[n]-A\cos(2\pi f_0 n+\Phi)\right]^2$$

Anulăm derivata acestei funcții

$$\frac{dJ(\Phi)}{d\Phi}=2\sum_{n=0}^{N-1}\left[x[n]-A\cos\left(2\pi f_0 n+\hat{\Phi}\right)\right]A\sin\left(2\pi f_0 n+\hat{\Phi}\right)=0$$

10

Rezolvarea ecuației, aplicând aproximările cunoscute, conduce la

$$\sum_{n=0}^{N-1} x[n]\sin\left(2\pi f_0 n + \hat{\Phi}\right) = \frac{A}{2}\sum_{n=0}^{N-1}\sin\left(4\pi f_0 n + 2\hat{\Phi}\right) \cong 0$$

cu condiția ca frecvența digitală să nu fie foarte apropiată de 0 sau 0.5. Dacă dezvoltăm membrul stâng al relației de mai sus obținem

$$\sum_{n=0}^{N-1} x[n]\sin\left(2\pi f_0 n + \hat{\Phi}\right) =$$

$$\sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n \cos\hat{\Phi} + \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n \sin\hat{\Phi} \cong 0$$

din care se determină estimatorul ML pentru faza inițială ca fiind

$$\hat{\Phi}_{ML} \cong -arctg\,\frac{\sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n}{\sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n} = -arctg\,\frac{T_2(\mathbf{x})}{T_1(\mathbf{x})}$$

11

relație în care am utilizat notațiile cunoscute deja

$$T_1(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]\cos 2\pi f_0 n;\quad T_2(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]\sin 2\pi f_0 n$$

În Cap.2 am determinat informația Fisher pentru faza inițială a sinusoidei

$$I(\Phi) \cong \frac{NA^2}{2\sigma^2}$$

Aplicând teorema enunțată putem afirma că repartiția asimptotică a estimatorului fazei este normală:

$$\hat{\Phi}_{ML} \overset{a}{\sim} \mathcal{N}\left(\Phi,\ \frac{2\sigma^2}{NA^2}\right)$$

Dispersia estimatorului ML al fazei tinde asimptotic spre

$$Disp\left\{\hat{\Phi}_{ML}\right\} \to \frac{2\sigma^2}{NA^2} = \frac{1}{N\eta};\quad \eta = \frac{A^2}{2\sigma^2} = SNR$$

12

Dacă introducem expresia semnalului de date în relația estimatorului ML al fazei inițiale obținem

$$\hat{\Phi}_{ML} \cong -arctg \frac{\sum_{n=0}^{N-1}\left[A\cos\left(2\pi f_0 n+\Phi\right)+w[n]\right]\sin 2\pi f_0 n}{\sum_{n=0}^{N-1}\left[A\sin\left(2\pi f_0 n+\Phi\right)+w[n]\right]\cos 2\pi f_0 n}$$

$$\cong -arctg \frac{-A\sum_{n=0}^{N-1}\sin\Phi + A\sum_{n=0}^{N-1}\sin\left(4\pi f_0 n+2\Phi\right)+2\sum_{n=0}^{N-1}w[n]\sin 2\pi f_0 n}{A\sum_{n=0}^{N-1}\cos\Phi + A\sum_{n=0}^{N-1}\cos\left(4\pi f_0 n+2\Phi\right)+2\sum_{n=0}^{N-1}w[n]\cos 2\pi f_0 n}$$

Dacă frecvența digitală nu este foarte aproape de 0 sau 0.5, termenii al doilea de la numărător și de la numitor pot fi neglijați, așa că rezultă

$$\hat{\Phi}_{ML} \cong -arctg \frac{-NA\sin\Phi + 2\sum_{n=0}^{N-1}w[n]\sin 2\pi f_0 n}{NA\cos\Phi + 2\sum_{n=0}^{N-1}w[n]\cos 2\pi f_0 n}$$

$$= arctg \frac{\sin\Phi - \frac{2}{NA}\sum_{n=0}^{N-1}w[n]\sin 2\pi f_0 n}{\cos\Phi + \frac{2}{NA}\sum_{n=0}^{N-1}w[n]\cos 2\pi f_0 n}$$

13

Estimarea cu plauzibilitate maximă în cazul transformării parametrilor

În unele cazuri nu ne interesează valoarea parametrului necunoscut, θ, de care depinde în mod nemijlocit semnalul de date, ci un parametru derivat, α=g(θ)

De exemplu, pentru datele, în care parametrul necunoscut este A

$$x[n] = A + w[n]; \quad n = 0,1,\ldots,N-1; \quad w[n] \sim \mathcal{N}\left(0, \quad \sigma^2\right)$$

dorim să estimăm

$$\alpha = e^A; \quad x[n] = \ln\alpha + w[n]$$

Cu datele **x** fixate, funcția de plauzibilitate dependentă de parametrul A este

$$p(\mathbf{x};A) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}; \quad A\in\mathbb{R}$$

Această funcție de plauzibilitate se poate se poate transforma ușor într-o funcție de plauzibilitate dependentă de parametrul α, deoarece, în acest caz funcția g(.) este biunivocă

$$p_\alpha(\mathbf{x};\alpha) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-\ln\alpha\right)^2\right\}$$

14

Se maximizează, în raport cu α, această funcție dependentă de α

$$\frac{d \ln p_\alpha(\mathbf{x};\alpha)}{d\alpha} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-\ln\hat{\alpha}\right)\frac{1}{\hat{\alpha}}$$

$$= \frac{1}{\hat{\alpha}\sigma^2}\left(\sum_{n=0}^{N-1}x[n]-N\ln\hat{\alpha}\right)=0$$

Din anularea parantezei rezultă logaritmul estimatorului ML pentru parametrul α

$$\ln\hat{\alpha}_{ML} = \frac{1}{N}\sum_{n=0}^{N-1}x[n]=\overline{x}$$

din care se obține estimatorul ML pentru parametrul α, obținut prin transformare

$$\hat{\alpha}_{ML} = e^{\overline{x}}$$

Se știe că media eșantion este estimatorul ML pentru componenta continuă, așa că putem scrie că

$$\hat{\alpha}_{ML} = e^{\hat{A}_{ML}} = g\left(\hat{A}_{ML}\right)$$

Relația de transformare a parametrilor este și relația de transformare a estimatorilor ML, cel puțin pentru transformările g(.) biunivoce

15

Vom da următoarea teoremă fără demonstrație.

Estimatorul ML pentru

$$\alpha = g(\theta)$$

unde θ parametrizează densitatea de probabilitate a datelor, p(**x**;θ), este dat de relația

$$\hat{\alpha}_{ML} = g\left(\hat{\theta}_{ML}\right)$$

Estimatorul $\hat{\theta}_{ML}$ se obține maximizând după θ funcția de plauzibilitate p(**x**;θ), cu **x** fixat

***

Vom considera un exemplu în care datele sunt un zgomot alb, gaussian, cu puterea necunoscută

$$x[n]=w[n];\quad n=0,1,\ldots,N-1;\quad w[n]\sim\mathcal{N}\left(0,\quad\sigma^2\right)$$

pentru care dorim să estimăm puterea în unități logaritmice

$$P = 10\log\sigma^2\ \ [\mathrm{dB}]$$

16

Deoarece zgomotul este alb, cu **x** fixat, obținem funcția de plauzibilitate

$$p\left(\mathbf{x};\sigma^2\right)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right\}$$

Se determină estimatorul ML pentru putere anulând derivata formei logaritmice

$$\frac{d\ln p\left(\mathbf{x};\sigma^2\right)}{d\sigma^2}=-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{n=0}^{N-1}x^2[n]=0$$

Din care se determină estimatorul ML pentru puterea zgomotului

$$\hat{\sigma}^2_{ML}=\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]$$

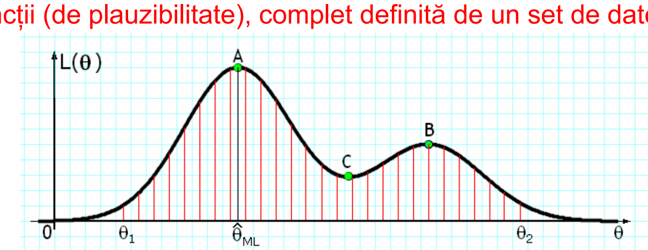
Substituim rezultatul în relația de transformare și obținem noul estimator ML

$$\hat{P}_{ML}=10\log\left(\alpha_{ML}\right)=10\log\left(\frac{1}{N}\sum_{n=0}^{N-1}x^2[n]\right)\quad[dB]$$

17

Determinarea numerică a estimării ML

Un avantaj al estimării cu plauzibilitate maximă este acela că putem determina o valoare estimată pentru fiecare set de date. În fapt, se caută maximul unei funcții (de plauzibilitate), complet definită de un set de date.



Dacă ştim că parametrul θ aparține unui interval precizat, aşa cum se arată în figură, putem recurge la o împărțire destul de fină a intervalului, printr-o grilă. Calculând valorile plauzibilității în punctele alese, putem stabili un argument ce maximizează , plauzibilitatea. Putem să recurgem şi la subîmpărțirea intervalului în care am găsit un maxim, în aşa fel încât localizarea punctului *A* (de maxim) să fie destul de bună. Există riscul să găsim un punct de maxim local, cum este *B*, în loc de maximul absolut, *A*. Este evident că pentru date noi, procedura de calcul trebuie reluată, deoarece valoarea estimatorului ML se schimbă.

18

Dacă domeniul valorilor posibile ale parametrului θ nu este un interval finit, căutarea maximului prin grilă pune serioase probleme de calcul. În astfel de cazuri se apelează la metode iterative de determinare a valorii argumentului ce maximizează plauzibilitatea. Dintre metodele utilizate vom prezenta, pe scurt, algoritmul Newton-Raphson. Modificarea datelor impune reluarea calculului, de fiecare dată. Trebuie specificat faptul că alegerea valorii inițiale influențează mult convergența algoritmului.

Pentru exemplificare, vom considera modelul de date x[n], în care zgomotul w[n] este alb, gaussian iar parametrul necunoscut, r, este pozitiv

$$x[n] = r^n + w[n]; \quad n = 0,1,\dots,N-1; \quad w[n] \sim \mathcal{N}\left(0, \quad \sigma^2\right)$$

Dacă în expresia densității de probabilitate (repartiție) fixăm vectorul de date **x**, obținem funcția de plauzibilitate, dependentă doar de parametrul necunoscut, r

$$p(\mathbf{x}; r) = \frac{1}{\left(2\pi\sigma^2\right)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n] - r^n\right)^2\right\}$$

Căutăm valoarea parametrului r care maximizează plauzibilitatea

19

Maximizarea plauzibilității este echivalentă cu minimizarea, după r, a funcției

$$J(r) = \sum_{n=0}^{N-1}\left(x[n] - r^n\right)^2$$

Estimatorul ML este soluția ecuației

$$\frac{dJ(r)}{dr} = \sum_{n=0}^{N-1} 2\left(x[n] - \hat{r}^n\right) n\hat{r}^{n-1} = 0$$

O astfel de ecuație nu poate fi rezolvată, în sensul stabilirii unei formule se calcul a soluției

***

Vom prezenta metode iterative de maximizare a plauzibilității logaritmice, ce caută un zerou al derivatei sale (o rădăcină):

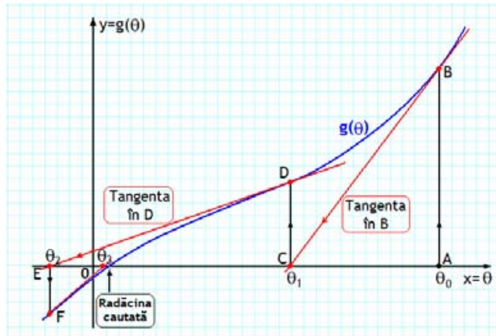
$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = 0$$

Soluția ecuației se determină în mod iterativ (prin mai mulți paşi), pornind de la o valoare inițială care se adoptă. Se notează cu g(θ) funcția

$$g(\theta) = \frac{d \ln p(\mathbf{x}; \theta)}{d\theta}$$

20

Algoritmul Newton-Raphson



Considerăm că valoarea de pornire în iterația pentru căutarea rădăcinii este θ_0. Ecuația tangentei la curba y=g(θ), dusă în punctul B (vezi figura) este

$$y - g(\theta_0) = \left.\frac{dg(\theta)}{d\theta}\right|_{\theta=\theta_0} (\theta - \theta_0)$$

Intersecția acestei tangente cu axa Ox, ce se obține punând y=0, definește o soluție mai bună pentru ecuația ce trebuie rezolvată. Ea este

$$\theta_1 = \theta_0 - \frac{g(\theta_0)}{g'(\theta_0)}; \quad g'(\theta_0) = \left.\frac{dg(\theta)}{d\theta}\right|_{\theta=\theta_0}$$

21

Se duce tangenta în punctul D și se obține o nouă soluție, corespunzătoare punctului E

$$\theta_2 = \theta_1 - \frac{g(\theta_1)}{g'(\theta_1)}$$

Valorile succesive ale soluțiilor pentru θ se determină cu relația de iterare

$$\theta_{k+1} = \theta_k - \frac{g(\theta_k)}{g'(\theta_k)}; \quad g'(\theta_k) = \left.\frac{dg(\theta)}{d\theta}\right|_{\theta=\theta_k} \quad k = 0,1,2,\ldots$$

Condiția de oprire a procesului iterativ de determinare a soluției este

$$\left|\theta_{k+1} - \theta_k\right| < \varepsilon$$

care afirmă că soluțiile succesive se află într-un interval cu lungimea mai mică decât eroarea maximă admisă ε

Substituind forma funcției g(θ) în formula de iterare, algoritmul Newton-Raphson pentru determinarea iterativă a estimatorului ML pentru θ devine

$$\theta_{k+1} = \theta_k - \left[\frac{d^2 \ln p(\mathbf{x};\theta)}{d\theta^2}\right]^{-1} \left.\frac{d \ln p(\mathbf{x};\theta)}{d\theta}\right|_{\theta=\theta_k} \quad k = 0,1,2,\ldots$$

22

Vom face câteva precizări privind aplicarea algoritmului Newton-Raphson:

i). Este posibil ca şirul de valori obținut prin iterație să nu conveargă. Acest lucru se întâmplă mai ales atunci când derivata a doua a plauzibilității logaritmice este de valoare apropiată de zero; drept consecință, termenul de corecție din relația (6.61) fluctuează mult, de la iterație la iterație.

ii). Chiar dacă şirul de valori obținut prin iterație converge, s-ar putea să nu ne conducă la un maxim global (punctul *A* din figură) ci la un maxim local (punctul *B* din figură), fie chiar într-un punct de minim (punctul *C* din figură). Pentru a evita cantonarea într-un maxim local este bine să lucrăm cu mai multe valori inițiale şi să reținem acea valoare care conduce la "cel mai mare maxim" al funcției de plauzibilitate. Alegerea valorii inițiale este deosebit de importantă într-o astfel de procedură iterativă.

23

Revenim la exemplul cu care am început. Avem, pentru derivatele plauzibilității logaritmice

$$\frac{d\ln p(\mathbf{x};r)}{dr} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-r^n\right)nr^{n-1}$$

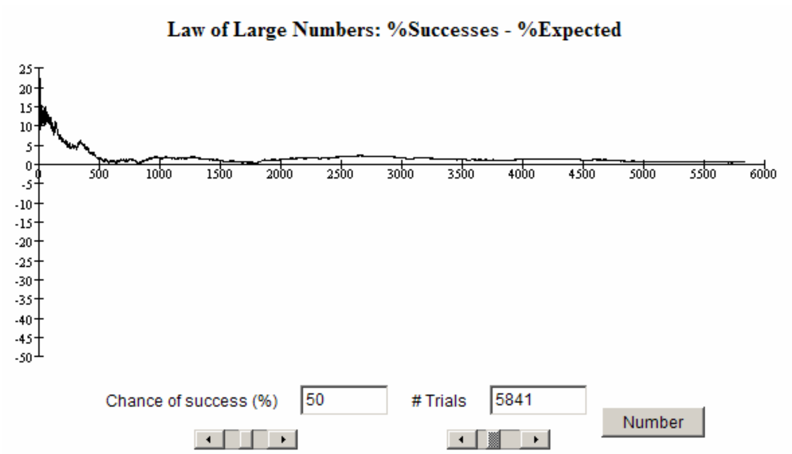
$$\frac{d^2\ln p(\mathbf{x};r)}{dr^2} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}nr^{n-2}\left[(n-1)x[n]-(2n-1)r^n\right]$$

Iterația pentru stabilirea estimatului ML pentru parametrul necunoscut, r, devine

$$r_{k+1} = r_k - \frac{\sum_{n=0}^{N-1}\left(x[n]-r_k^n\right)nr_k^{n-1}}{\sum_{n=0}^{N-1}nr_k^{n-2}\left[(n-1)x[n]-(2n-1)r_k^n\right]}$$

24

O versiune a metodei Newton-Raphson face uz de legea numerelor mari, care afirmă că pentru eşantioane IID media aritmetică este o aproximare bună a mediei statistice, dacă numărul de eşantioane mediate este mare



25

Conform legii numerelor mari, dacă N este suficient de mare, putem scrie, pentru cazul a N eşantione de date x[n], statistic independente şi identic distribuite (IID) că

$$\frac{d^2 \ln p(\mathbf{x};\theta)}{d\theta^2} = \sum_{n=0}^{N-1} \frac{d^2 \ln p(x[n];\theta)}{d\theta^2}$$

$$= N \underbrace{\frac{1}{N} \sum_{n=0}^{N-1} \frac{d^2 \ln p(x[n];\theta)}{d\theta^2}}_{\downarrow}$$

$$\cong N \overbrace{E\left\{\frac{d^2 \ln p(\mathbf{x};\theta)}{d\theta^2}\right\}}^{\downarrow}$$

$$= -Ni(\theta) = -I(\theta)$$

Pentru cazul în care N este suficient de mare şi eşantioanele de date sunt IID, formula de iterare a algoritmului Newton-Raphson devine

$$\theta_{k+1} = \theta_k + I^{-1}(\theta_k) \left.\frac{d \ln p(\mathbf{x};\theta)}{d\theta}\right|_{\theta = \theta_k}; \qquad k = 0,1,2,\ldots$$

Metoda descrisă prin această relaţie este cunoscută şi sub denumirea de "Method of Scoring"

26

Revenim (a doua oară) la exemplul de început. Pentru cazul în speță, informația Fisher este

$$I(\theta) = -E\left\{\frac{d^2 \ln p(\mathbf{x};r)}{dr^2}\right\}$$

$$= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1} n r^{n-2}\left[(n-1)E\{x[n]\} - (2n-1)r^n\right]$$

$$= -\frac{1}{\sigma^2}\sum_{n=0}^{N-1} n r^{n-2}\left[(n-1)r^n - (2n-1)r^n\right]$$

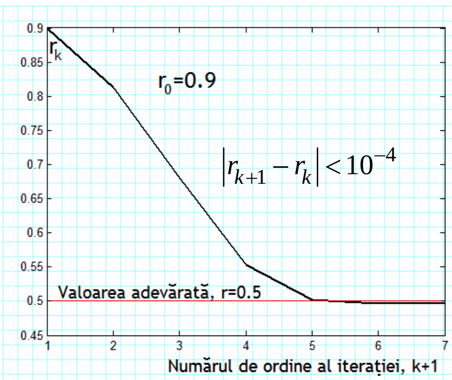
$$= \frac{1}{\sigma^2}\sum_{n=0}^{N-1} n^2 r^{2n-2}$$

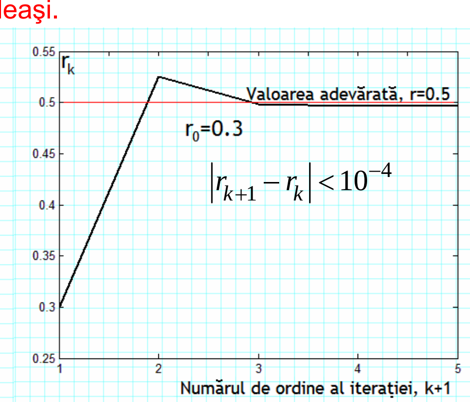
Dacă o substituim în formula de iterație “method of scoring” obținem

$$r_{k+1} = r_k + \frac{\sum_{n=0}^{N-1}\left(x[n] - r_k^n\right) n r_k^{n-1}}{\sum_{n=0}^{N-1} n^2 r_k^{2n-2}}$$

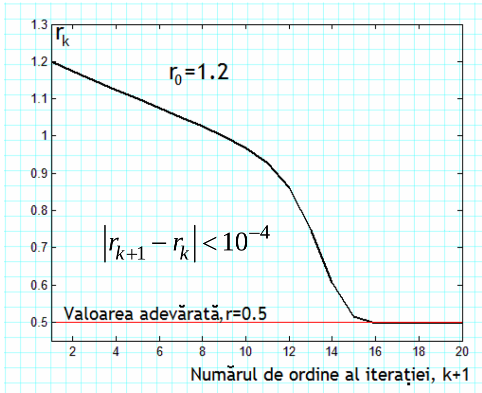
27

Pentru cazul valorii reale r=0.5 şi pentru dispersia de 0.01, s-a efectuat o simulare pentru estimarea MLE cu relația ce implementeză “method of scoring”, pentru o realizare a datelor cu lungimea N=50. Condiția de oprire a simulării a fost ε=0.0001. În figură se pot urmări valorile r succesive, dacă valoarea inițială a fost 0.9 . După numai 6 iterații se constată că valorile succesive au primele patru zecimale egale, aşa că valoarea estimatului ML este 0.4969. Dacă valoarea inițială e luată 0.3 se constată că după numai patru iterații se obține condiția de oprire. Valoarea estimării este aceeaşi deoarece datele sunt aceleaşi.





Dacă însă valoarea inițială este 1.2, condiția de oprire se atinge doar după 17 iterații



. Se poate observa că valoarea inițială influențează mult numărul de iterații necesare pentru calculul estimării ML.

29

O extindere pentru cazul parametrului vector

Reluăm problema determinarii a doi parametri necunoscuți, dacă modelul de semnal corespunde unei componente continue, afectată de un zgomot alb, gaussian

$$x[n] = A + w[n]; \quad n = 0,1,\ldots,N-1; \quad w[n] \sim \mathcal{N}\left(0, \quad \sigma^2\right)$$

Vectorul parametrilor necunoscuți este

$$\boldsymbol{\theta} = \begin{bmatrix} A & \sigma^2 \end{bmatrix}^T$$

Cu datele **x** fixate, se determină plauzibilitatea. Pentru a stabili estimatorul ML, se anulează derivata plauzibilității logaritmice, în raport cu A

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial A} = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(x[n] - A\right) = 0$$

şi se obține

$$\hat{A}_{ML} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] = \overline{x}$$

Se anulează apoi derivata în raport cu al doilea parametru

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=0}^{N-1}\left(x[n] - A\right)^2$$

30

Conform celor prezentate anterior, valoarea pentru A se înlocuieşte cu estimarea sa ML

$$A \leftarrow \hat{A}_{ML} = \overline{x}$$

şi se obține ecuația

$$-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=0}^{N-1}\left(x[n]-\overline{x}\right)^2 = 0$$

Găsim al doilea estimator ML de forma

$$\overset{\wedge}{\sigma^2_{ML}} = \frac{1}{N}\sum_{n=0}^{N-1}\left(x[n]-\overline{x}\right)^2$$

Cei doi estimatori scalari se pot grupa într-un estimator ML vector

$$\hat{\boldsymbol{\theta}}_{ML} = \begin{bmatrix} \overline{x} \\ \frac{1}{N}\sum_{n=0}^{N-1}\left(x[n]-\overline{x}\right)^2 \end{bmatrix}$$

31

Dăm, fără demonstrație, teorema

Dacă densitatea de probabilitate a datelor, **x**, p(**x**; θ**)**, ce depinde de parametrul vector θ**,** necunoscut, satisface condiția de regularitate

$$E\left\{\frac{\partial p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} = \mathbf{0}$$

atunci estimatorul ML pentru parametrul vector necunoscut, θ, ce maximizează plauzibilitatea, este repartizat asimptotic, conform cu

$$\hat{\boldsymbol{\theta}}_{ML} \overset{a}{\sim} \mathcal{N}\left(\boldsymbol{\theta},\ \ \mathbf{I}^{-1}(\boldsymbol{\theta})\right)$$

unde **I**(θ) este matricea de informație Fisher, cu elementele evaluate la valorile adevărate ale componentelor vectorului θ

***

32

Reluăm problema determinarii a doi parametri necunoscuți, dacă modelul de semnal corespunde unei componente continue, afectată de un zgomot alb, gaussian

$$x[n] = A + w[n];\ n = 0,1,\dots,N-1;\ w[n] \sim \mathcal{N}(0,\ \sigma^2);\ \boldsymbol{\theta} = \begin{bmatrix} A & \sigma^2 \end{bmatrix}^T$$

pentru care tocmai am stabilit estimatorii ML. Cunoaştem repartiția mediei

$$\bar{\mathbf{x}} \sim \mathcal{N}\left(A,\ \frac{\sigma^2}{N}\right)$$

Mai ştim că statistica T are o repartiție hi-pătrat, şi este statistic independentă de media eşantion

$$T = \sum_{n=0}^{N-1}\left(\frac{x[n]-\bar{x}}{\sigma}\right)^2 \sim \chi^2_{N-1}$$

Literatura afirmă că repartiția hi-pătrat converge spre una normală

$$\chi^2_N \underset{N\to\infty}{\longrightarrow} \mathcal{N}(N,\ 2N)$$

Avem deci pentru statistica T o repartiție asimptotică normală

$$T \sim \chi^2_{N-1} \overset{a}{\sim} \mathcal{N}\left(N-1,\ 2(N-1)\right)$$

Forma estimatorului ML pentru dispersia datelor a fost stabilit de noi și este

$$\hat{\sigma}^2_{ML} = \frac{1}{N}\sum_{n=0}^{N-1}\left(x[n]-\bar{x}\right)^2 = \frac{\sigma^2}{N}\sum_{n=0}^{N-1}\left(\frac{x[n]-\bar{x}}{\sigma}\right)^2 = \frac{\sigma^2}{N}T$$

Media estimatorului se determină cu relația

$$E\left\{\hat{\sigma}^2_{ML}\right\} = \frac{\sigma^2}{N}E\{T\} = \frac{\sigma^2}{N}(N-1)$$

iar dispersia are forma

$$Disp\left\{\hat{\sigma}^2_{ML}\right\} = \frac{\sigma^4}{N^2}Disp\{T\} = \frac{\sigma^4}{N^2}2(N-1)$$

Se poate da acum repartiția asimptotică a estimatorului ML pentru dispersie

$$\hat{\sigma}^2_{ML} \overset{a}{\sim} \mathcal{N}\left(\frac{N-1}{N}\sigma^2,\ 2\frac{N-1}{N^2}\sigma^4\right)$$

Cele două componente ale vectorului estimator sunt repartizate normal și sunt statistic independente. Componentele vectorului sunt deci și mutual gaussiene. Caracterizarea sa statistică se realizează, în consecință, doar prin vectorul mediilor și prin matricea de covarianță

Vectorul mediilor demonstrează o comportare asimptotică de neabatere

$$E\left\{\hat{\boldsymbol{\theta}}_{ML}\right\} = \begin{bmatrix} E\{\bar{x}\} \\ E\left\{\hat{\sigma}^2_{ML}\right\} \end{bmatrix} = \begin{bmatrix} A \\ \frac{N}{N-1}\sigma^2 \end{bmatrix} \xrightarrow[N\to\infty]{} \begin{bmatrix} A \\ \sigma^2 \end{bmatrix} = \boldsymbol{\theta}$$

Matricea de covarianță are pe diagonala principală dispersiile. Celelalte două elemente sunt nule, ca urmare a lipsei corelației între cei doi estimatori ML scalari

$$\mathbf{C}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & 2\frac{N-1}{N^2}\sigma^4 \end{bmatrix} \xrightarrow[N\to\infty]{} \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix} = \mathbf{I}^{-1}(\boldsymbol{\theta})$$

Se vede imediat că matricea de covarianță tinde asimptotic spre inversa matricei Fisher, stabilită de noi anterior

Am verificat, prin calcul direct, faptul că teorema teorema privind repartiția asimptotică a estimatorului vector ML se aplică în exemplul de față:

$$\hat{\boldsymbol{\theta}}_{ML} \overset{a}{\sim} \mathcal{N}\left(\boldsymbol{\theta},\ \ \mathbf{I}^{-1}(\boldsymbol{\theta})\right)$$

35

Teorema de invarianță a estimatorului ML vector

Estimatorul de plauzibilitate maximă (ML) pentru vectorul

$$\boldsymbol{\alpha}_{r\times 1}$$

obținut din vectorul

$$\boldsymbol{\theta}_{p\times 1}$$

prin funcția inversabilă α=g(θ) este

$$\hat{\boldsymbol{\alpha}}_{ML} = g\left(\hat{\boldsymbol{\theta}}_{ML}\right)$$

Dacă funcția α=g(θ) nu este inversabilă, atunci estimatorul ML pentru α asigură maximizarea funcției de plauzibilitate transformată

$$p_T(\mathbf{x};\boldsymbol{\alpha}) = \max_{\{\boldsymbol{\theta}:\ \boldsymbol{\alpha}=\mathbf{g}(\boldsymbol{\theta})\}} p(\mathbf{x};\boldsymbol{\theta})$$

***

36

Vom exemplifica cu modelul de semnal componentă continuă afectată de un zgomot alb, gaussian, pentru care nu se cunosc A și puterea zgomotului

$$x[n] = A + w[n]; \quad n = 0,1,\dots,N-1; \quad w[n] \sim \mathcal{N}\left(0, \quad \sigma^2\right)$$

dar ne interesează doar raportul semnal/zgomot, SNR

$$\alpha = \frac{A^2}{\sigma^2}$$

Având în vedere cele expuse avem pentru estimatorul SNR

$$\hat{\alpha}_{ML} = \frac{\left(\hat{A}_{ML}\right)^2}{\hat{\sigma}^2_{ML}} = \frac{(\bar{x})^2}{\frac{1}{N}\sum_{n=0}^{N-1}\left(x[n]-\bar{x}\right)^2}$$

Vom aborda acum cazul modelului gaussian generalizat, în care zgomotul este gaussian, avand media nulă şi matricea de covarianță **C**(θ)

$$\mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \quad \mathbf{C}(\boldsymbol{\theta})\right)$$

Repartiția datelor **x** este normală

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \quad \mathbf{C}(\boldsymbol{\theta})\right)$$

37

Se poate arăta că pentru repartiția normală de mai înainte, este valabilă regula de derivare a plauzibilității

$$\frac{\partial p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} tr\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k}\right\} + \frac{\partial \boldsymbol{\mu}^T(\boldsymbol{\theta})}{\partial \theta_k}\mathbf{C}^{-1}(\boldsymbol{\theta})\left[\mathbf{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\right]$$
$$-\frac{1}{2}\left[\mathbf{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\right]^T \frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k}\left[\mathbf{x}-\boldsymbol{\mu}(\boldsymbol{\theta})\right]; \; k = 1,2,\dots,p$$

Modelul liniar generalizat este descris de relația:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

$\mathbf{H}_{Nxp}$ fiind matricea de observare

După ce datele **x** au fost fixate, expresia funcției de plauzibilitate este

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\right\}$$

Maximizarea plauzibilității implică minimizarea funcției:

$$J(\boldsymbol{\theta}) = (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

38

Putem să recurgem direct la maximizarea plauzibilității, ținând seama de regula de derivare anterior expusă. Avem, pentru medie

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$$

și deci

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_k} = \frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \theta_k}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta});\quad k = 1, 2, \ldots, p$$

Dacă scriem în mod explicit cele p relații, acestea pot fi puse sub formă matriceală

$$\begin{bmatrix} \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix} = \begin{bmatrix} \frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \theta_1} \\ \frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \theta_2} \\ \vdots \\ \frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \theta_p} \end{bmatrix}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

39

sau, cu notațiile matriceale uzuale

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

Cunoaștem regula de derivare

$$\frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} = \mathbf{H}^T$$

care, după aplicare, ne dă

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{H}^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

Egalăm derivata cu **0** și obținem estimatorul ML pentru vectorul θ

$$\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\hat{\boldsymbol{\theta}}_{ML} = \mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

cu expresia

$$\hat{\boldsymbol{\theta}}_{ML} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

Rezultă că, pentru modelul liniar generalizat, estimatorul de plauzibilitate maximă este şi estimator MVU şi este şi eficient

$$\hat{\boldsymbol{\theta}}_{ML} = \hat{\boldsymbol{\theta}}_{MVU}$$

40

Estimatorul are matricea de covarianță de forma

$$\mathbf{C}_{\hat{\theta}} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}$$

Teorema de optimalitate a estimatorului ML pentru modelul liniar. Dacă datele observate $\mathbf{x}$ pot fi descrise prin modelul liniar:

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

în care $\mathbf{H}_{N\times p}$ este o matrice de observare cunoscută, cu $N > p$ şi de rang p, $\boldsymbol{\theta}_{p\times 1}$ este vectorul necunoscut ce trebuie estimat iar:

$$\mathbf{w}_{N\times 1} \sim \mathcal{N}\left(\mathbf{0},\ \ \mathbf{C}\right)$$

atunci estimatorul ML pentru $\boldsymbol{\theta}$ este:

$$\hat{\boldsymbol{\theta}}_{ML} = \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}$$

$\hat{\boldsymbol{\theta}}_{ML}$ este şi estimator MVU şi este şi eficient, în sensul că îşi atinge limita inferioară Cramer-Rao. Repartiția estimatorului este normală:

$$\hat{\boldsymbol{\theta}}_{ML} = \hat{\boldsymbol{\theta}}_{MVU} \sim \mathcal{N}\left(\hat{\boldsymbol{\theta}}\ \ \left(\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H}\right)^{-1}\right)$$

***

Vom considera că într-o problemă de estimare, în care datele **x** conduc la funcția de plauzibilitate p(**x**; θ), există un estimator eficient

$$\hat{\boldsymbol{\theta}}_{ef}$$

Fiind eficient, el satisface teorema Cramer-Rao, adică

$$\frac{\partial \ln p\left(\mathbf{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}} = \mathbf{I}\left(\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_{ef} - \boldsymbol{\theta}\right)$$

Estimatorul ML satisface condiția de maxim a plauzibilității logaritmice

$$\left.\frac{\partial \ln p\left(\mathbf{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{ML}} = \mathbf{0}$$

Substituind valoarea estimatorului ML în expresia teoremei Cramer-Rao obținem

$$\left.\frac{\partial \ln p\left(\mathbf{x};\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{ML}} = \mathbf{I}\left(\hat{\boldsymbol{\theta}}_{ML}\right)\left(\hat{\boldsymbol{\theta}}_{ef} - \hat{\boldsymbol{\theta}}_{ML}\right) = \mathbf{0}$$

Rezultă concluzia

$$\hat{\boldsymbol{\theta}}_{ML} = \hat{\boldsymbol{\theta}}_{ef}$$

Concluzia se poate enunța ca teoremă.

Dacă într-o problemă de estimare există un estimator eficient, atunci procedura de determinare a estimatorului ML îl va produce, cei doi estimatori fiind egali.

Estimarea asimptotică cu plauzibilitate maximă

În literatura de specialitate se arată că, pentru N suficient de mare, adică asimptotic, plauzibilitatea logaritmică se poate aproxima cu relația

$$\ln p(\mathbf{x};\boldsymbol{\theta}) \cong -\frac{N}{2}\ln 2\pi - \frac{N}{2}\int_{-1/2}^{1/2}\left[\ln P_{xx}(f) + \frac{I(f)}{P_{xx}(f)}\right]df$$

relație în care $P_{xx}(f)$ este densitatea spectrală de putere a datelor iar I(f) este periodograma datelor **x**

$$I(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j2\pi fn}\right|^2 = \frac{1}{N}\left|X(f)\right|^2$$

Dependența plauzibilității de parametrul vector **θ** are loc prin $P_{xx}(f)$. Derivata plauzibilității logaritmice după componenta θ_i se calculează cu relația

$$\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_i} \cong -\frac{N}{2}\int_{-1/2}^{1/2}\left[\frac{1}{P_{xx}(f)} - \frac{I(f)}{P_{xx}(f)}\right]\frac{\partial P_{xx}(f)}{\partial \theta_i}df; \quad i = 1, 2, \dots, p$$

Pentru stabilirea celor p componente ale estimatorui vector ML, este necesară rezolvarea sistemului de ecuații

$$\int_{-1/2}^{1/2}\left[\frac{1}{P_{xx}(f)} - \frac{I(f)}{P_{xx}(f)}\right]\frac{\partial P_{xx}(f)}{\partial \theta_i}df = 0; \quad i = 1, 2, \dots, p$$

43

Câteva exemple din domeniul prelucrării semnalelor

Estimarea distanței în RADAR și în SONAR

Am determinat, pentru RADAR, densitatea de repartiție a datelor. Dacă în expresia acesteia se consideră datele fixate, se obține funcția de plauzibilitate

$$p(\mathbf{x};n_0) = \prod_{n=0}^{n_0-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2[n]}{2\sigma^2}\right\}\cdot$$

$$\prod_{n=n_0}^{n_0+M-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\left(x[n]-s[n-n_0]\right)^2}{2\sigma^2}\right\}\cdot$$

$$\prod_{n=n_0+M}^{N-1}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{x^2[n]}{2\sigma^2}\right\}$$

în care parametrul necunoscut este durata de propagare a semnalului de sondaj

$$n_0\Delta = \tau_0$$

44

Expresia plauzibilității poate fi adusă la o formă mai simplă

$$p(\mathbf{x};n_0)=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}x^2[n]\right\}\cdot$$

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=n_0}^{n_0+M-1}\left(-2x[n]s[n-n_0]+s^2[n-n_0]\right)\right\}$$

în care primul factor nu depinde de parametrul necunoscut. Pentru a găsi un estimator ML este necesar să maximizăm al doilea factor, ceeace echivalează cu minimizarea expresiei

$$\sum_{n=n_0}^{n_0+M-1}\left(-2x[n]s[n-n_0]+s^2[n-n_0]\right)=$$

$$-2\sum_{n=n_0}^{n_0+M-1}x[n]s[n-n_0]+\sum_{n=n_0}^{n_0+M-1}s^2[n-n_0]$$

Al doilea termen este o constantă, egală cu energia semnalului de sondare

$$\sum_{n=n_0}^{n_0+M-1}s^2[n-n_0]=\sum_{m=0}^{M-1}s^2[m]=\mathcal{E}_S$$

Minimizarea expresiei analizate este echivalentă cu maximizarea funcției 45

$$J(n_0)=\sum_{n=n_0}^{n_0+M-1}x[n]s[n-n_0]$$

Valoarea care maximizează funcția J este estimatorul ML a timpului

$$\hat{n}_{0ML}$$

Cu această estimare se determină estimatorul ML pentru distanța RADAR-țintă

$$\hat{R}_{ML}=\frac{c\Delta}{2}\hat{n}_{0ML}$$

Estimarea parametrilor semnalului sinusoidal

Vom relua problema estimării celor trei parametrii ai semnalului sinusoidal, al cărui model este

$$x[n]=A\cos(2\pi f_0 n+\Phi)+w[n];\ \ n=0,1,\ldots,N-1;\ w[n]\sim\mathcal{N}(0,\ \ \sigma^2)$$

Parametrul vector luat]n considerare are forma

$$\boldsymbol{\theta}=\begin{bmatrix}A & f_0 & \Phi\end{bmatrix}^T$$

Cu datele **x** fixate, funcția de plauzibilitate, dependentă de vectorul θ este

$$p(\mathbf{x};\boldsymbol{\theta})=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left[x[n]-A\cos(2\pi f_0 n+\Phi)\right]^2\right\}$$ 46

Pentru a maximiza plauzibilitatea este necesar să minimizăm, după θ, expresia

$$J(\boldsymbol{\theta}) = J(A, f_0, \Phi) = \sum_{n=0}^{N-1}\left[x[n] - A\cos(2\pi f_0 n + \Phi)\right]^2$$

$$= \sum_{n=0}^{N-1}\left(x[n] - A\cos\Phi\cos 2\pi f_0 n + A\sin\Phi\sin 2\pi f_0 n\right)^2$$

Cu notațiile

$$\alpha_1 = A\cos\Phi; \quad \alpha_2 = -A\sin\Phi$$

forma funcției de minimizat, J, devine

$$J(A, f_0, \Phi) = \sum_{n=0}^{N-1}\left(x[n] - \alpha_1\cos 2\pi f_0 n - \alpha_2\sin 2\pi f_0 n\right)^2$$

Sunt evidente relațiile

$$A = \sqrt{\alpha_1^2 + \alpha_2^2}; \quad \Phi = arctg\,\frac{-\alpha_2}{\alpha_1}$$

47

Definim vectorii **c** (cosinus) și **s** (sinus)

$$\mathbf{c} = \left[1 \quad \cos 2\pi f_0 \quad \cos 2\pi f_0\cdot 2 \quad \cdots \quad \cos 2\pi f_0\cdot(N-1)\right]^T$$

$$\mathbf{s} = \left[0 \quad \sin 2\pi f_0 \quad \sin 2\pi f_0\cdot 2 \quad \cdots \quad \sin 2\pi f_0\cdot(N-1)\right]^T$$

și construim forma

$$\mathbf{x} - \alpha_1\mathbf{c} - \alpha_2\mathbf{s} = \begin{bmatrix} x[0] - \alpha_1\cdot 1 - \alpha_2\cdot 0 \\ x[1] - \alpha_1\cos 2\pi f_0 - \alpha_2\sin 2\pi f_0 \\ x[2] - \alpha_1\cos 2\pi f_0\cdot 2 - \alpha_2\sin 2\pi f_0\cdot 2 \\ \vdots \\ x[N-1] - \alpha_1\cos 2\pi f_0(N-1) - \alpha_2\sin 2\pi f_0(N-1) \end{bmatrix}$$

Prin calcul direct se verifică faptul că funcția J poate fi pusă sub forma

$$J(A, f_0, \Phi) = J(\alpha_1, \alpha_2, f_0)$$
$$= (\mathbf{x} - \alpha_1\mathbf{c} - \alpha_2\mathbf{s})^T(\mathbf{x} - \alpha_1\mathbf{c} - \alpha_2\mathbf{s})$$

48

Facem notațiile pentru matricea **H** și vectorul α

$$\mathbf{H}=\begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \cos 2\pi f_0\cdot 2 & \sin 2\pi f_0\cdot 2 \\ \vdots & \vdots \\ \cos 2\pi f_0(N-1) & \sin 2\pi f_0(N-1)\end{bmatrix}=[\mathbf{c}\quad \mathbf{s}]\ ;\ \boldsymbol{\alpha}=\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}$$

Cu acestea, funcția de minimizat, J, devine

$$J(\alpha_1,\alpha_2,f_0)=(\mathbf{x}-\mathbf{H}\boldsymbol{\alpha})^T(\mathbf{x}-\mathbf{H}\boldsymbol{\alpha})$$

Vom minimiza mai întâi după α şi apoi după frecvența digitală. Am minimizat mai devreme o altă funcție

$$J(\boldsymbol{\theta})=(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

Dacă punem

$$\mathbf{C}^{-1}=\mathbf{I}_u$$

stabilim, prin similitudine că

$$\hat{\boldsymbol{\alpha}}_{ML}=(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

Cu valorile determinate, rezultă pentru J o formă ce va fi minimizată după f

$$J(\hat{\alpha}_1,\hat{\alpha}_2,f_0)=(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\alpha}})^T(\mathbf{x}-\mathbf{H}\hat{\boldsymbol{\alpha}})$$

$$=\left(\mathbf{x}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}\right)^T\left(\mathbf{x}-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}\right)$$

$$=\mathbf{x}^T\left[\mathbf{I}_u-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right]\left[\mathbf{I}_u-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right]\mathbf{x}$$

49

Matricea factor din parantezele drepte este evident simetrică şi idempotentă

$$\left[\mathbf{I}_u-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right]\left[\mathbf{I}_u-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right]=$$

$$\mathbf{I}_u-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T+\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$$

$$=\mathbf{I}_u-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$$

Cu acestea funcția de minimizat, J, ia forma

$$J(\hat{\alpha}_1,\hat{\alpha}_2,f_0)=\mathbf{x}^T\left[\mathbf{I}_u-\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\right]\mathbf{x}$$

$$=\mathbf{x}^T\mathbf{x}-\mathbf{x}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

Minimizarea funcției J este echivalentă cu maximizarea termenului al doilea, primul fiind o constantă

$$J_1(f_0)=\mathbf{x}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$$

$$=\mathbf{x}^T[\mathbf{c}\quad \mathbf{s}]\left(\begin{bmatrix}\mathbf{c}^T\\ \mathbf{s}^T\end{bmatrix}[\mathbf{c}\quad \mathbf{s}]\right)^{-1}\begin{bmatrix}\mathbf{c}^T\\ \mathbf{s}^T\end{bmatrix}\mathbf{x}$$

$$=\mathbf{x}^T[\mathbf{c}\quad \mathbf{s}]\begin{bmatrix}\mathbf{c}^T\mathbf{c} & \mathbf{c}^T\mathbf{s}\\ \mathbf{s}^T\mathbf{c} & \mathbf{s}^T\mathbf{s}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{c}^T\\ \mathbf{s}^T\end{bmatrix}\mathbf{x}$$

50

Se stabilesc relațiile aproximative, dacă frecvența digitală nu este foarte aproape de 0 sau 0.5

$$\mathbf{c}^T\mathbf{s}=\mathbf{s}^T\mathbf{c}=\sum_{n=0}^{N-1}\cos 2\pi f_0 n\sin 2\pi f_0 n=\frac{1}{2}\sum_{n=0}^{N-1}\sin 4\pi f_0 n\cong 0$$

$$\mathbf{c}^T\mathbf{c}=\sum_{n=0}^{N-1}\cos^2 2\pi f_0 n=\frac{1}{2}\sum_{n=0}^{N-1}\left(1+\cos 4\pi f_0 n\right)\cong\frac{N}{2}$$

$$\mathbf{s}^T\mathbf{s}=\sum_{n=0}^{N-1}\sin^2 2\pi f_0 n=\frac{1}{2}\sum_{n=0}^{N-1}\left(1-\cos 4\pi f_0 n\right)\cong\frac{N}{2}$$

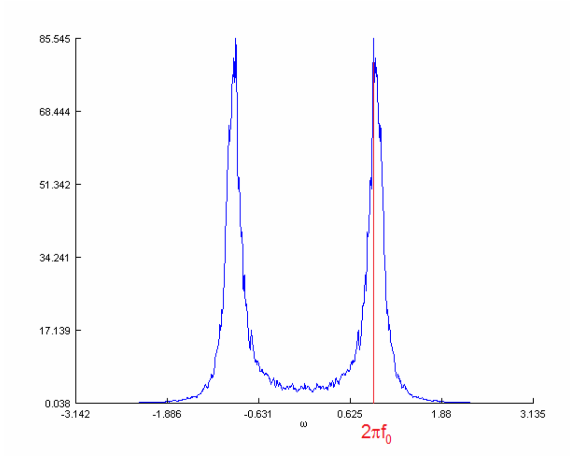
51

Prin calcul direct arătăm că funcția de maximizat este dublul periodogramei. E suficient să determinăm frecvența la care apare maximul perodogramei pentru a avea estimatorul ML pentru frecvența digitală

$$\begin{aligned}
J_1(f_0)&=\mathbf{x}^T[\mathbf{c}\quad\mathbf{s}]\begin{bmatrix}\frac{N}{2} & 0\\ 0 & \frac{N}{2}\end{bmatrix}^{-1}\mathbf{x}\\
&=\frac{2}{N}\left(\begin{bmatrix}\mathbf{c}^T\\ \mathbf{s}^T\end{bmatrix}\mathbf{x}\right)^T\left(\begin{bmatrix}\mathbf{c}^T\\ \mathbf{s}^T\end{bmatrix}\mathbf{x}\right)\\
&=\frac{2}{N}\begin{bmatrix}\mathbf{c}^T\mathbf{x}\\ \mathbf{s}^T\mathbf{x}\end{bmatrix}^T\begin{bmatrix}\mathbf{c}^T\mathbf{x}\\ \mathbf{s}^T\mathbf{x}\end{bmatrix}\\
&=\frac{2}{N}\left[\left(\mathbf{c}^T\mathbf{x}\right)^2+\left(\mathbf{s}^T\mathbf{x}\right)^2\right]\\
&=\frac{2}{N}\left[\left(\sum_{n=0}^{N-1}x[n]\cos 2\pi f_0 n\right)^2+\left(\sum_{n=0}^{N-1}x[n]\sin 2\pi f_0 n\right)^2\right]\\
&=\frac{2}{N}\left|\left(\sum_{n=0}^{N-1}x[n]e^{-j2\pi f_0 n}\right)\right|^2=\frac{2}{N}\left|X(f_0)\right|^2
\end{aligned}$$

52

$$I(f)=\frac{1}{N}\left|X(f)\right|^2 \qquad J_1(f_0)=2I(f_0)$$



$$\hat{f}_{0ML}=\arg\max_{f_0} I(f)$$

Frevența digitală la care apare maximul periodogramei calculată pentru datele **x** din realizarea curentă, determină estimatorul ML pentru frecvență.

53

Având estimatorul pentru frecvență, se determină estimatorii ML pentru componentele vectorului α

$$\hat{\boldsymbol{\alpha}}_{ML}=\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{x}=\frac{2}{N}\mathbf{I}_u\begin{bmatrix}\mathbf{c}^T\mathbf{x}\\ \mathbf{s}^T\mathbf{x}\end{bmatrix}$$

$$\hat{\boldsymbol{\alpha}}_{ML}=\begin{bmatrix}\hat{\alpha}_{1ML}\\ \hat{\alpha}_{2ML}\end{bmatrix}=\begin{bmatrix}\frac{2}{N}\sum_{n=0}^{N-1}x[n]\cos 2\pi\hat{f}_0 n\\ \frac{2}{N}\sum_{n=0}^{N-1}x[n]\sin 2\pi\hat{f}_0 n\end{bmatrix}$$

Se stabilește estimatorul ML pentru amplitudinea sinusoidei

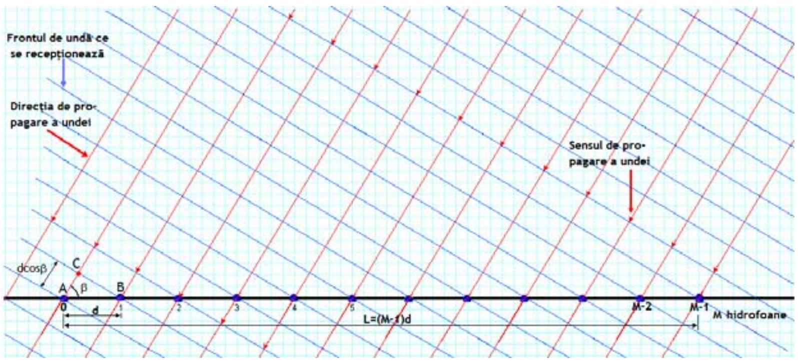
$$\hat{A}_{ML}=\sqrt{\hat{\alpha}_1^2+\hat{\alpha}_2^2}=\frac{2}{N}\left|\sum_{n=0}^{N-1}x[n]e^{-j2\pi\hat{f}_0 n}\right|=\frac{2}{N}\left|X\left(\hat{f}_0\right)\right|$$

54

În final, se stabilește estimatorul ML pentru faza inițială a sinusoidei

$$\hat{\Phi}_{ML} = -arctg\frac{\hat{\alpha}_2}{\hat{\alpha}_1} = -arctg\frac{\sum_{n=0}^{N-1} x[n]\cos 2\pi \hat{f}_0 n}{\sum_{n=0}^{N-1} x[n]\sin 2\pi \hat{f}_0 n}$$

Determinarea direcției din care sosesc undele sonore în SONAR



55

Direcția din care vine un semnal sonor este legată de frecvența spațială care are formula

$$f_S = f_0\frac{d}{c}\cos\beta \qquad [f_0] = Hz$$

Dacă sunt de estimat ML amplitudinea, frecvența digitală și faza inițială, suntem în situația studiată chiar mai înainte. Vom presupune că frecvența analogică, geometria de dispunere a hidrofoanelor precum și viteza de propagare a sunetului în apă sunt cunoscute. Pentru a estima frecvența spațială, cu datele **x** construim periodograma

$$I(f_S) = \frac{1}{M}\left|\sum_{n=0}^{N-1} x[n]e^{-j2\pi f_S n}\right|^2$$

și determinăm, din poziția vârfului ei, estimatorul

$$\hat{f}_{S,ML}$$

Se determină apoi estimatorul ML al unghiului β

$$\hat{\beta}_{ML} = \arccos\frac{\hat{f}_{S,ML}c}{f_0 d}$$

56