Terminologia de procesării digitale a semnalelor

Semnale şi grafice

În literatura de limba engleză acest domeniu este referit prin **DSP - Digital Signal Processing**.

Un semnal este o descriere a modului în care un parametru este dependent de un alt parametru. De exemplu, cel mai des întalnit tip de semnal în electronica analogică este o tensiune care depinde de timp. Deoarece ambii parametrii pot acoperii un interval continuu de valori, acest tip de semnal il vom numi **semnal continuu**. În contrast, trecerea acestui semnal printr-un convertor analog digital forţează quantizarea ambilor parametrii. Spre exemplu, dacă conversia se face cu un convertor de 12 biti la o rată de de 1000 de eşantioane pe secundă, tensiunea va putea lua doar 4096 valori posibile, iar timpul este definit doar la incremente de o milisecundă. Semnalele formate din parametrii care sunt cuantizaţi în acest mod se numesc **semnale discrete** sau **semnale digitale**. Sunt de asemenea posibile semnale la care unul dintre parametrii este continuu iar celălalt este discret. Întrucât acest gen de semnale sunt rar folosite ele n-au primit un nume explicit. În aceste cazuri tipul celor două semnale trebuie precizat explicit.

Figura 2.1 prezintă două semnale aşa cum ar putea rezulta dintr-un sistem de achiziţii de date.

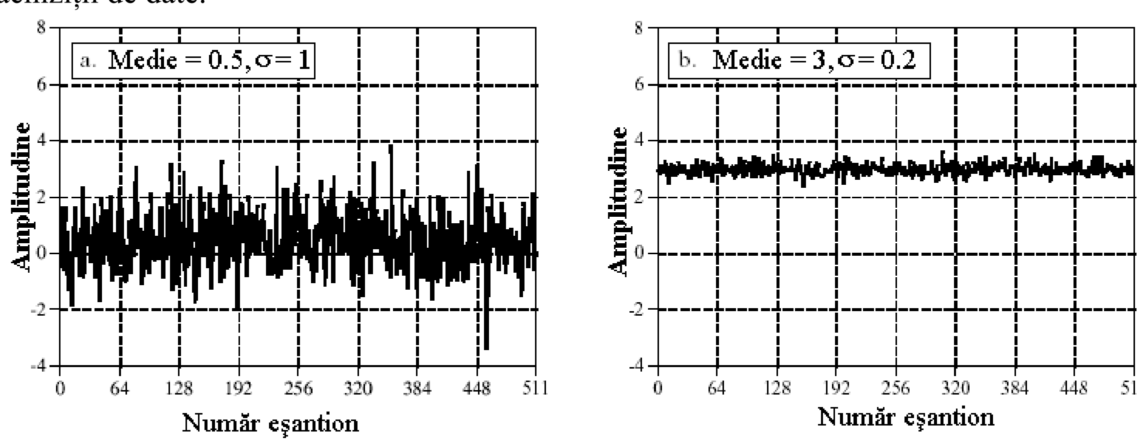


Fig. 2.1.

Axa verticală poate reprezenta tensiunea, intensitatea luminii, presiunea sunetului, sau oricare altă mărime despre care deţinem această informaţie. Deoarece nu ştim ce reprezintă în acest caz particular, îi vom da numele generic: amplitudine. Acest parametru este apelat şi cu alte denumiri precum: axa y, variabila dependentă, gama sau ordonata.

Axa orizontală reprezintă celălalt parametru al semnalului, care are şi el diferite denumiri precum: axa x, variabila dependentă, domeniu sau abscisa. Timpul este cel mai des întâlnit parametru pe axa orizontală a unui semnal achiziţionat. Cu toate acestea se pot întâlni şi alte mărimi fizice pe această axă. Spre exemplu, geofizicienii pot achiziţiona semnale referitoare la densitatea unei roci la distanţe echidistante pe suprafaţa pământului. Pentru a păstra lucrurile la un grad ridicat de generalitate noi vom denumi axa orizontală: număr eşantion. Dacă acesta ar fi fost un semnal continuu aceeastă axa ar fi fost denumită corespunzător: timp, distanţă, x, etc.

Cei doi parametrii care formează un semnal sunt de regulă neinterschimbabili. Parametrul de pe axa y (variabila dependentă) se spune că este o funcţie de parametrul de pe axa x (variabila independentă). Cu alte cuvinte, variabila independentă descrie cum şi când fiecare eşantion este prelevat, în timp ce variabila dependentă reprezintă valoarea măsurată efectiv. O caracteristică a acestor reprezentări funcţionale este că unei valori de pe axa x îi este întotdeauna asociată o valoare pe axa y. Reciproca nu este adevărată întotdeauna.

Trebuie acordată o atenţie particulară cuvântului domeniu, un foarte larg utilizat termen în DSP. Spre exemplu, un semnal care utilizează drept variabilă independentă timpul, se spune că aparţine **domeniului timp** (time domain). O altă clasă de semnale usuale în DSP utilizează frecvenţa ca variabilă independentă, rezultând termenul de domeniu frecvenţă (frequency domain). Analog, semnalele care utilizează distanţa drept parametru independent aparţin domeniului spaţial (spatial domain). În concluzie tipul parametrului de pe axa orizontală reprezintă domeniul acelui semnal.

Deşi semnalele din Fig. 2.1 sunt discrete, ele sunt reprezentate în această figură prin linii continue. Aceasta deoarecă există prea multe eşantioane apropiate pentru a fi percepute ca separate. În graficele unor semnale scurte, să zicem 100 de eşantioane, marcherii eşantioanelor se pot vedea separat. Linii continue care să conecteze aceşti marcheri pot fi sau nu trasate între marcherii consecutivi, după cum autorul graficului doreşte sau nu acest lucru.

Variabila N este larg utilizată in DSP pentru a reprezenta numărul total de eşantioane ale unui semnal. Spre exemplu, pentru semnalul din Fig. 2.1. N = 512. Pentru a păstra datele în ordine, fiecărui eşantion îi este atribuit un număr, numit uneori index. Aceste numere sunt cele ce apar de-a lungul axei orizontale. Două notaţii sunt de regulă folosite pentru atribuirea numerelor de eşantion. În prima notaţie indexul eşantioanele merge de la 1 la N (de exemplu de la 1 la 512). A doua notaţie, indexul merge de la 0 la N-1 (de ex., de la 0 la 511). Matematicienii o folosesc de regulă prima notaţie, în timp ce în DSP este folosită mai mult a doua notaţie. Această dublă notaţie poate crea uneori probleme. De aceea un bun obicei atunci când sunt parcurse diferite materiale este identificarea încă de la început a notaţiei pe care o foloseşte autorul respectivului material. În literatura de limbă engleză pentru desemnarea celor două notaţii de folosesc termenii: *one based index*, respectiv *zero based index*.

Media şi abaterea standard

Media (mean), notată cel mai adesea cu μ, este numele prescurtat a ceea ce statisticienii numesc *valoare medie* (average value) a valorilor luate de o *variabilă aleatoare*, definită cu ajutorul *mediei aritmetice* a tuturor eşantioanelor disponibile

$$\mu = \mu_N = \frac{1}{N}\sum_{k=1}^{N} x_k .$$

Pentru simplificarea scrierii vom utiliza bara superioară pentru desemnarea mediilor aritmetice ale unor funcţii de variabila aleatoare

$$\overline{f(x)}^N = \frac{1}{N}\sum_{k=1}^{N} f(x_k) .$$

În acest sens, $\mu_N = \overline{x}^N$.

În electronică, *media* unui semnal este numită *componenta continuă* a semnalului (Direct Current DC). Similar, AC (alternating current) se referă la modul în care fluctuează un semnal în jurul valorii sale medii. Dacă semnalul are o formă de undă simplă, precum unda sinusoidală sau pătrată, pentru a măsura excursia acestui semnal se foloseşte amplitudinea vârf la vârf (peak-to-peak amplitude).

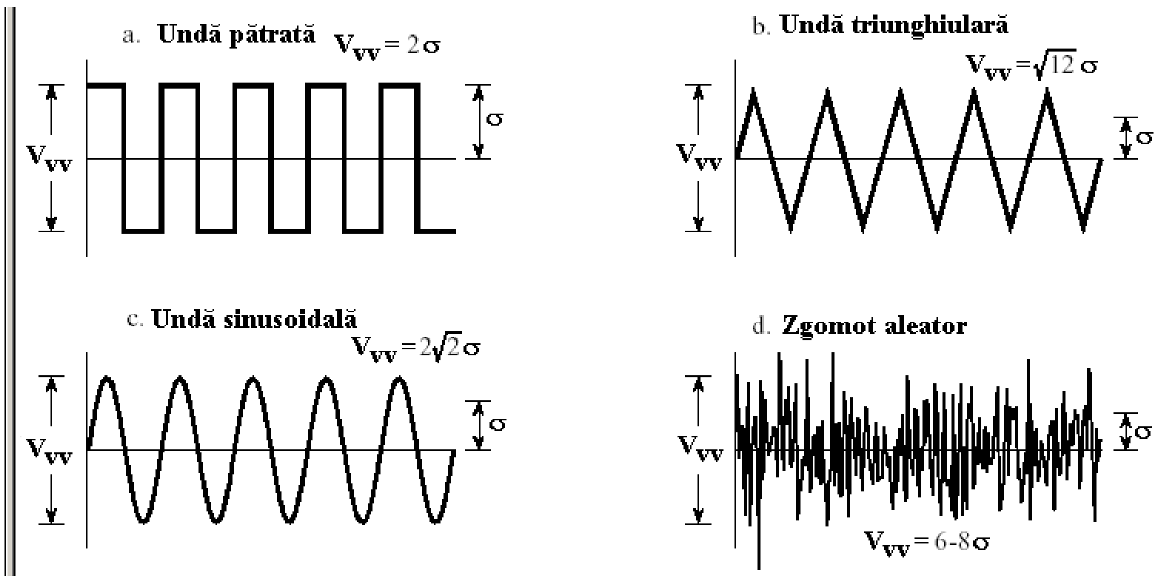


Fig. 2.2. Raportul dintre amplitudinea vârf la vârf şi abaterea standard pentru câteva forme de undă comune. Pentru unda pătrată, acest raport este 2; pentru unda triunghiulară este $\sqrt{12} \cong 3.46$; pentru unda sinusoidală este $2\sqrt{2} \cong 2.83$, în timp ce pentru o undă aleatoare nu are un raport exact. În funcție de distribuția acestui semnal acest poate lua diferite valori de regulă între 6 şi 8.

În practică, calculul mediei se poate face chiar în timpul colectării datelor (fie printr-un proces de achiziție a datelor, fie de calcul, al doilea caz fiind specific modelării). Rata cu care sunt obținute datele poate fi mare astfel încât, în scurt timp, volumul datelor ajunge impresionant şi stocarea lor devine o problemă chiar şi pentru calculatoare performante din acest punct de vedere. De aceea se preferă estimarea dinamică a mediilor (running averaging) fără însă a se reține toate aceste valori. Aceste lucru este posibil dacă se remarcă relația de recurență ce leagă două medii consecutive

$$m\mu_m + x_{m+1} = (m+1)\mu_{m+1}.$$

Această relație arată că, pentru calculul mediei curente, este necesară memorarea mediei anterioare, a numărului de eşantioane şi cunoaşterea ultimei valori luată de variabila aleatoare.

Abaterea medie (average deviation). Evident, expresia $|x_k - \mu|$ arată cât de mult se abate (deviates) valoarea x_k de la medie. Media aritmetică a acestor abateri reprezintă, cum este de aşteptat, abaterea medie. În multe cazuri, parametrul important nu este abaterea de la medie, ci puterea acetei abateri. Spre exemplu, atunci când semnale de zgomot aleator afectează un circuit electronic, puterea zgomotului resultat este suma puterilor semnalelor individuale. Prin urmare mărimi legate de amplidudinea semnalelor componente caracterizează mai slab astfel de semnale. De aceea în locul abaterii medii este folosită abaterea standard atunci când se doreşte o informaţie medie legată de amplitudinea unui semnal.

Abaterea standard (standard deviation) sau *eroarea pătratică medie* este rădăcina pătrată a mediei aritmetice a pătratelor abaterilor de la medie.

$$\sigma_N = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(x_k-\mu)^2}\ .$$

Abaterea standard este o măsură a distanţei până la care variabila aleatoare (semnalul) fluctuează faţă de medie deci a dispersiei. Observăm în expresia de mai sus o neconcordanţă, în loc de N s-a luat $N-1$. Aceasta apare din următorul motiv: la momentul în care noi estimăm abaterea standard nu cunoaştem media adevărată μ, ci numai o estimare a sa în acel moment. Aceasta conduce la o subevaluare a abaterii standard. De aceea, pentru valori mici ale lui N este bine să se împartă la $N-1$ în loc de N pentru a compensa această diferenţă. Evident la valori mari ale lui N diferenţa între cele două estimări nu mai contează.

Estimarea dinamică a abaterii standard beneficiază şi ea de o relaţie ce permite estimarea dinamică a sa. Într-adevăr, suma de sub radical se poate scrie

$$\sum_{k=1}^{N}(x_k-\mu_N)^2 = \sum_{k=1}^{N}x_k^2 - 2\mu_N\sum_{k=1}^{N}x_k + m\mu_N^2,$$

cum însă $\sum_{k=1}^{N}x_k = m\mu_N$ rezultă

$$\sigma_N^2 = \frac{1}{N-1}\left[\sum_{k=1}^{N}x_k^2 - \frac{1}{N}\left(\sum_{k=1}^{N}x_k\right)^2\right] = \frac{1}{N-1}\left[\text{suma pătratelor} - \frac{\text{suma}^2}{N}\right].$$

Prin urmare, calculul abaterii standard curente necesită memorarea a trei numere: sumei, a sumei pătratelor, a numărului eşantioanelor şi cunoaşterea ultimei valori luată de variabila aleatoare.

Mărimea σ^2 este cunoscută în statistică sub numele de *varianţă* sau *dispersie*. În teoria semnalelor, varianţa reprezintă puterea acestor fluctuaţii de la valoare medie a semnalului. Într-adevăr, puterea medie disipată de un semnal de tensiune $U(t)$ a cărui *componentă alternativă* (alternative current AC) este $u(t)$ pe un rezistor R este

$$\overline{P}^T = \frac{1}{T}\int_0^T \frac{u^2(t)}{R}dt\ .$$

Însă media din formula precedentă se calculează pe baza eşantioanelor discrete colectate cu perioada h pe un interval temporal de lungime $T = Nh$, spre exemplu. Dacă se consideră un rezistor de valoare unitară $R = 1$ şi se notează eşantioanele tensiunii $U(t)$ cu x_k formula de mai înainte conduce imediat la definiţia dată varianţei.

Folosind notațiile prescurtate ale mediilor varianța se poate scrie sub forma

$$\sigma^2 = \overline{x^2} - \left(\overline{x}\right)^2 .$$

O altă mărime legată de varianță este *rădăcina mediei pătratelor* (root-mean-square rms) definită prin

$$rms(x) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} x_k^2} = \sqrt{\overline{x^2}} .$$

Prin definiție, abaterea standard măsoară doar parte AC a unui semnal, în timp ce rms măsoară ambele componente (AC şi DC). Evident, rms coincide cu abaterea standard atunci când semnalul are medie nulă.

În mecanică există o altă semnificație interesantă a abaterii standard legată de noțiunea de *rază de girație*. Prin definiție, raza de girație este distanța fața de axa de rotație unde ar trebui concentrată toată masa corpului pentru a da acelaşi moment de inerție față de acea axă [HRI 82, p. 108]. Astfel, pentru un corp constituit din N puncte materiale de mase m_k plasate la distanțe r_k de axa de rotație, raza de girație este

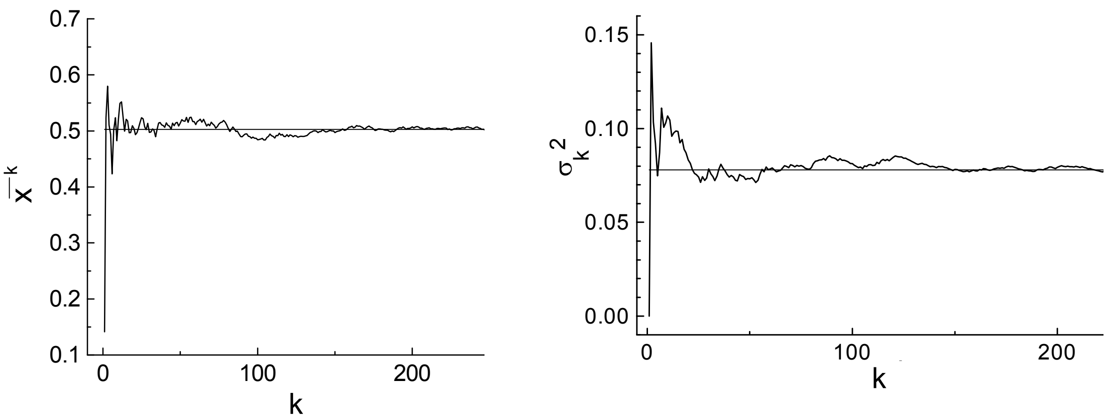
$$R_g = \sqrt{\frac{1}{m}\sum_{k=1}^{N} m_k r_k^2} ,$$

unde $m = \sum_{k=1}^{N} m_k$ este masa totală.

Considerăm acum corpul constituit din N mase elementare toate egale cu unitatea, plasate pe o singură dreaptă. Într-un reper liniar, definit pe suportul punctelor materiale, pozițiile punctelor materiale sunt precizate prin coordonatele x_k, iar poziția centrului de masă prin coordonata μ. Alegem drept axă de rotație o dreaptă perpendiculară pe suportul punctelor materiale ce trece prin centrul de masă al corpului. Dacă distanțele particulelor la această axă sunt $r_k = |x_k - \mu|$ rezultă imediat că raza de girație coincide cu definiția abaterii standard.

Atunci când setul de date provine dintr-un *proces de măsurare* asupra valori unei mărimi fizice, *media* reprezintă o estimare a valorii mărimii măsurate, iar *abaterea standard* caracterizează procesul de măsurare calificând *precizia* cu s-au efectuat măsurările.

În figura 2.3a este reprezentată evoluția *mediei dinamice* către media adevărată în cazul unei variabile cu repartiție uniformă (U) în intervalul [0,1), iar în figura 7b evoluția varianței dinamice către varianța aceleiaşi variabile.



(a) **(b)**

Fig. 2.3. Evoluția: (a) *mediei dinamice* către media unei variabile cu repartiție uniformă în intervalul [0,1); (b) *varianței dinamice* către varianța aceleiaşi variabile.

În anumite situații, media caracterizează rezultatul unei măsurări (în esență valoarea măsurată a unei mărimi), în timp ce abaterea standard caracterizează zgomotul şi alte elemente perturbatoare ce pot afecta procesul de măsurare. În aceste situații, abaterea standard nu este importantă prin ea înseşi ci prin comparație cu media. Așa a fost întrodus termenul de raport semnal zgomot (signal-to-noise-ratio SNR).

În *teoria semnalelor* avem de a face cu altă situație: un semnal lent variabil în timp, peste care este suprapus un zgomot, a cărei variație este mai rapidă decât a semnalului. În această situație, medierea (actul de măsurare propriuzis) se face pe un interval de timp intermediar, adică suficient de lung pentru a atenua zgomotul însă suficient de scurt pentru a surprinde variația lentă a semnalului. Acestui domeniu îi este propriu termenul de *raport semnal zgomot*, care este egal cu media împărțită la abaterea standard

$$\mathrm{SNR}(x) = \frac{\overline{x}}{\sigma_x}.$$

Simultan cu SNR coexistă termenul *coeficient de variație* (coefficient of variation CV) definit prin

$$\mathrm{CV}(x) = \frac{\sigma_x}{\overline{x}}.$$

În mod clar cei doi termeni califică acelaşi lucru. Un proces de măsurare corect presupune o valoare mare a SNR (sau o valoare mică a CV).

Evoluția *mărimilor statistice dinamice* către *valorile adevărate*

Staționaritate. Un fenomen (proces) este considerat staționar dacă **proprietățile statistice ale sale sunt aceleaşi oricare ar fi originea timpului**. Considerăm necesare câteva comentarii referitor la această definiție. În fapt, toate informațiile despre un sistem (fenomen) pe care experimentul la furnizează cercetătorului sunt limitate la timpul cât durează experimentul. În esență, un experiment înseamnă colectarea mai multor seturi de date referitoare la una sau mai multe mărimi fizice. Ipoteza staționarității presupune identitatea tuturor proprietăților statistice ale seturilor de date colectate într-un experiment.

Prin definiție, un proces care nu satisface ipoteza anterioară este *nestaționar*.

Realitatea experimentală este crudă. Dintr-o multitudine de cauze, nu prea există procese staționare. Rezultă că un proces staționar este un proces ideal din categoria multor concepte ideale din fizică (vezi de exemplu conceptul de ciclul Carnot).

Deosebim două surse primare de nestaționaritate. Prima se referă chiar la nestaționaritatea fenomenului de studiat, iar a doua la nestaționaritatea procesului de măsurare.

Nestaționaritatea fenomenului de studiat (ca şi a procesului de măsurare) poate fi *proprie* (sau intrinsecă) sau *indusă* de interacția cu mediul exterior.

De aceea, o bună parte din efortul unui experimentator este focalizat către *aducerea interacției* sistemului investigat cu exteriorul *la un nivel cât mai scăzut* prin ecranări, stabilizări, izolări, etc., scopul acestui demers fiind *izolarea fenomenului de studiat şi a procesului de măsurare de influențele mediului exterior*.

Odată realizat acest deziderat rămân în studiu fenomenele nestaționare intrinseci.

O situație experimentală întâlnită des este următoarea. Din investigații prealabile se cunosc constantele de timp ale proceselor nestaționare. Există două situații experimentale limită. (i) Pentru procese nestaționare cu constantă mare de timp experimentatorul va căuta să obțină o cantitate cât mai mare de date într-un timp cât mai mic în comparație cu constanta de timp a procesului nestaționar. (ii) Dacă nestaționaritățile au constantă de timp mică ele se vor trata ca fluctuații ale mărimilor de interes. În această situație intervalul de timp de colectare a datelor va fi mult mai mare decât perioada "fluctuațiilor" procesului nestaționar. Scopul ambelor demersuri este aducerea procesului de măsurare cât mai aproape de un proces staționar. Un proces experimentat în astfel de condiții se va numi evident *cuasistaționar*.

În concluzie, ideal pentru experimentator este obținerea unor *date staționare* din indiferent ce conjunctură experimentală s-ar afla. Nici un efort nu va fi prea mic pentru atingerea acestui ideal.

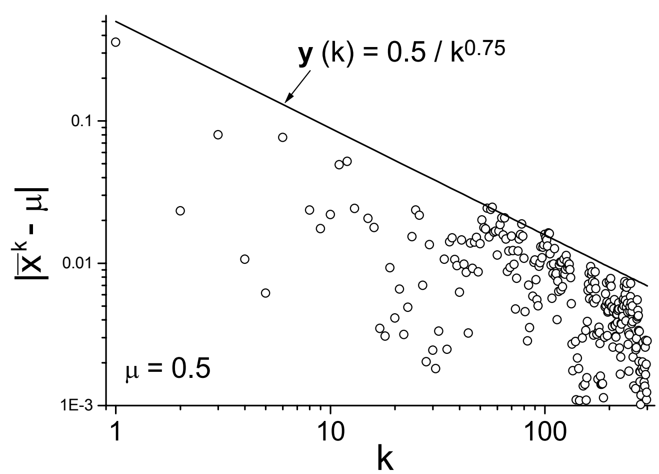


Fig. 2.4.

O altă situație experimentală este următoarea. La orice moment al achiziției de date pot fi calculate mărimile statistice cu datele deja disponibile (noi le-am numit generic *medii dinamice* deoarece ele depind de timp, sau ceea ce e totuna, de numărul de eşantioane). Experimentatorul doreşte să ştie cât de aproape este *media sa* de *media adevărată*. Pentru aceasta este nevoie de un *estimator* al *abaterii maxime de la media adevărată*.

Spre ilustrare, reprezentăm în figura 2.4 evoluţia abaterii de la media adevărată pentru media reprezentată în figura 2.3a. Ce constatăm? Abaterea de la media adevărată devine din ce în ce mai mică pe măsură ce creşte numărul eşantioanelor. La marginea superioară a datelor reprezentate în figura 2.4 este reprezentată funcţia

$$y(k) = \frac{0.5}{k^{0.75}}.$$

Această funcţie reprezintă estimatorul găsit de noi, în acest caz particular, pentru abaterea maximă de la mărimea adevărată.

Desigur, fiecărei medii i se poate asocia câte un estimator al abaterii maxime de la medie. Pornind de la aceste observaţii se poate defini *eroarea tipică* (typical error TE). În esenţă ne aşteptăm ca după un număr oarecare de măsurări media obţinută să difere de media adevărată a procelului studiat. Mărimea ce caracterizează această eroare se numeşte eroare tipică. În particular, pentru semnale aleatoare, eroarea tipică TE, dintre media calculată folosind datele a N măsurători şi media adevărată a procesului studiat este dată de:

$$TE = \frac{\sigma}{N^{1/2}}.$$

Proprietăţi ale *mediei* şi *dispersiei*.

Evident, *media unei constante* este chiar constanta

$$\overline{c} = c.$$

Orice *constantă multiplicativă* poate fi scoasă factor în procesul de mediere

$$\overline{cx} = c\overline{x}$$

Media sumei a două variabile aleatoare este egală cu suma mediilor

$$\overline{x+y} = \overline{x} + \overline{y}.$$

Media produsului a două variabile aleatoare este

$$\overline{x \cdot y} = \overline{x} \cdot \overline{y} + \operatorname{cov}(x, y).$$

Atunci când cele două variabile sunt independente rezultă imediat

$$\overline{x \cdot y} = \overline{x} \cdot \overline{y}.$$

Dispersia unei constante este nulă.

$$\sigma_c^2 = \overline{c^2} - \left(\overline{c}\right)^2 = 0.$$

Un *factor constant* iese de sub *operatorul dispersie* ridicat la pătrat.

$$\sigma_{cx}^2 = \overline{(cx)^2} - \left(\overline{cx}\right)^2 = c^2\sigma_x^2.$$

Notă. Statisticienii notează *operatorul de mediere* aplicat unei variabile aleatoare x cu $M(x)$, ceea ce în notațiile noastre înseamnă

$$M(x) = \bar{x},$$

în timp ce operatorul dispersie este definit prin

$$D(x) = M\left([x - M(x)]^2\right).$$

Noi am preferat notația cu bară datorită simplității scrierii.

Dispersia unei sume de două variabile aleatoare este egală cu suma dispersiilor celor două variabile plus dublul covarianței lor.

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 + 2\operatorname{cov}(x, y).$$

Deoarece

$$\operatorname{cov}(ax, y) = \operatorname{cov}(x, ay) = a\operatorname{cov}(x, y),$$

rezultă generalizarea

$$\sigma_{ax+by}^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\operatorname{cov}(x, y)$$

Corelație, convoluție, covarianță

Coeficientul de corelație. Fie două seturi de date provenind de la două variabile aleatorii $\{X\}$, $\{Y\}$. În analiza zgomotelor se doreşte, spre exemplu, să se identifice gradul în care două semnale sunt afectate de aceeaşi sursă de zgomot.

O metodă de comparație se poate obține presupunând existența unei legături de proporționalitate între cele două seturi de date, de exemplu de forma $y \cong \alpha x$. *Metoda celor mai mici pătrate* (least squares method) permite estimarea acestei constante de proporționalitate prin minimalizarea sumei

$$S = \sum_{k=1}^{N} (y_k - \alpha x_k)^2 .$$

Minimul acestei sume în raport cu parametrul α se caută rezolvând ecuația $\partial S / \partial\alpha = 0$. Rezultă astfel

$$\alpha = \left(\sum_{k=1}^{N} x_k y_k\right) / \left(\sum_{k=1}^{N} x_k^2\right).$$

Dacă substituim această valoare a parametrului α în suma precedentă obținem

$$S_{min} = (1 - \rho^2)\sum_{k=1}^{N} y_k^2 ,$$

unde

$$\rho = \left(\sum_{k=1}^{N} x_k y_k\right) / \sqrt{\sum_{k=1}^{N} x_k^2 \sum_{k=1}^{N} y_k^2} ,$$

se numeşte *coeficient de corelaţie*. *Teorema Schwarz-Buniakovski*, din matematica elementară, arată că

$$\left(\sum_{k=1}^{N} x_k y_k\right)^2 \le \left(\sum_{k=1}^{N} x_k^2\right)\left(\sum_{k=1}^{N} y_k^2\right),$$

şi prin urmare $0 \le |\rho| \le 1$. Sintetic, *coeficientul de corelaţie* este

$$\rho = \frac{\overline{xy}}{\sqrt{\overline{x^2}\,\overline{y^2}}}.$$

Atunci când $\rho = 0$ se spune că cele două *procese* sunt *necorelate* sau *independente*, iar când $\rho = 1$ corelaţia este totală ($\alpha = 1$ şi prin urmare $x_k = y_k$, $k = 1,2,\ldots,N$). Iată de ce, o măsură a gradului de proporţionalitate a celor două variabile o reprezintă acest coeficient de corelaţie. Mai remarcăm irelevanţa semnului acestui coeficient.

Covarianţa. Prin definiţie covarianţa a două seturi de valori luate de două variabile aleatoare este

$$\mathrm{cov}(x,y) = \frac{1}{N}\sum_{k=1}^{N}\left(x_k - \overline{x}\right)\left(y_k - \overline{y}\right) = \overline{\left(x - \overline{x}\right)\left(y - \overline{y}\right)}.$$

Pe baza celor prezentate până acum, remarcăm că

$$\left[\mathrm{cov}(x,y)\right]^2 \le \sigma_x^2 \sigma_y^2.$$

Evident, covarianţa califică gradul de corelare al abaterilor de la medie.

În electronică, acea parte a semnalului care reprezintă abaterea de la medie se numeşte *componentă alternativă* (alternative current AC).

Precizia şi acurateţea.

Precizia şi *acurateţea* sunt termeni utilizaţi pentru a descrie procese (sisteme şi metode) care măsoară, estimează sau prezic *valoarea adevărată* pe care o mărime fizică o are. După cum se cunoaşte, procesul de măsurare furnizează o *valoare măsurată* pe care o dorim cât mai apropiată de valoarea adevărată. Prin urmare acurateţea şi precizia sunt moduri de a descrie erorile care există între aceste două valori.

Acurateţea specifică abaterea valorii măsurate de la valoarea adevărată a mărimii măsurate.

Precizia reprezintă o măsură a dispersiei valorilor mărimii măsurate obţinute în procesul de măsurare. În *procesul de măsurare* două sisteme sunt puse în interacţie: (i) sistemul asupra căruia se efectuează măsurarea (furnizorul mărimii de măsurat) şi (ii) sistemul care efectuează măsurarea.

Dispersia valorilor mărimii măsurate şi prin urmare *precizia* pot fi calificate statistic în mai multe moduri. În prezent, cel mai utilizat calificator al preciziei îl constituie *abaterea standard*.

Deoarece valorile individuale obținute prin măsurare pot diferi substanțial una de alta pot fi utilizate drept calificatori ai preciziei de măsurare *abaterea standard*, SNR sau CV, toate trei calificând dispersia valorilor măsurate (Fig. 2.5).

Pentru ilustrarea acestor definiții, am reprezentat în figura 2.5. histograma valorilor luate de media dinamică reprezentată în figura 2.3a.

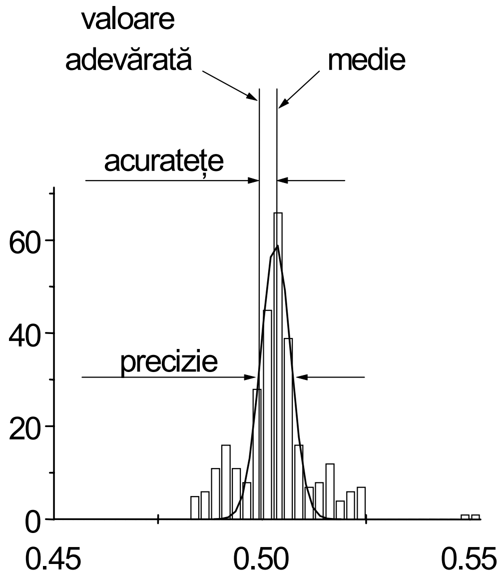


Fig. 2.5. Acuratețe, precizie, valoare adevărată.

Ponderi, funcții densitate de repartiție.

Într-o *histogramă* a unei variabile aleatoare discrete o anumită valoare x_k apare n_k ori. În această situație media variabilei se calculează cu

$$\bar{x} = \frac{\sum_{k=1}^{N} x_k n_k}{N} = \sum_{k=1}^{N} x_k p_k \ ,$$

unde N reprezintă numărul de valori pe care le poate lua variabila aleatoare, iar raportul $p_k = n_k / N$ se numeşte ponderea cu care valoarea x_k apare în spectrul de valori al variabilei. Un exemplu de variabilă cu spectru finit este răspunsul unui convertor analog numerice **CNA** (Digital Analog Converter **DAC**) de 8 biți. Spectrul de valori al unui astfel de convertor acoperă 256 numere întregi în gama 0..255.

În general, o variabilă aleatoare este asociată unui fenomen (proces) staționar. În această situație ponderile p_k tind, pentru un număr foarte mare de eşantioane ale variabilei aleatoare, către numere reale unice P_k, numite *probabilități* de apariție a valorilor x_k.

$$P_k = \lim_{N\to\infty}\left(\frac{n_k}{N}\right).$$

Atunci când variabila aleatoare este cu variație continuă media se calculează analog

$$\bar{x} = \int_I x f(x) dx \,,$$

unde I este intervalul în care ia valori variabila aleatoare, iar funcția $f(x)$ se numeşte complet *funcția densitate de repartiție/probabilitate* (probability density/distribution function **pdf**). Evident, **pdf** păstrează semnificația ponderii şi de aceea uneori mai este numită *funcție pondere*.

Dacă notăm cu $P(X < x)$ probabilitatea ca variabila X să ia valori mai mici decât o valoare x precizată funcția definită în acest fel

$$F(x) = P(X < x),$$

se numeşte *funcție de repartiție* (cumulative distribution function **cdf**). Pe baza acestei funcții este uşor acum să definim **pdf** prin

$$f(x) = \frac{dF(x)}{dx} \,,$$

având semnificația imediată: dacă $dF(x)$ este probabilitatea ca variabila aleatoare să ia valori în intervalul $(x, x + dx)$, atunci $f(x)$ este probabilitatea pe unitatea de lungime.

Teorema limită centrală

Semnalele obținute din procese aleatorii au de regulă o foarte frumoasă funcție de distribuție. Aceasta este *distribuția normală*, sau *distribuția Gauss*, sau pe scurt *Gaussiană*. Motivul pentru care se întâmplă acest lucru este explicat de o teoremă foarte importantă: *Teorema limită centrală*.

Teorema limită centrală, în forma sa cea mai simplă afirmă că o sumă de numere aleatoare devine *normal distribuită* pe măsură ce creşte numărul de numere aleatoare care sunt adunate. De remarcat că teorema nu specifică necesitatea ca numerele însumate să aparțină unei *legi de distribuție* anume.

Distribuția normală (The normal distribution)

O variabilă aleatoare, având media nulă, are o *distribuție/repartiție normală* dacă funcția sa densitate este

$$f(x) = A e^{-\beta x^2} \,,$$

unde A este *constanta de normare*, iar β un parametru specific.

Constanta de normare se calculează imediat impunând ca

$$\int_{-\infty}^{\infty} A e^{-\beta x^2} dx = 1 \,.$$

Există o foarte frumoasă şi utilă demonstrație a formulei de calcul a integralei de mai sus. Se înmulțesc două astfel de integrale, una în variabilă x iar cealaltă în variabilă y, apoi se trece în coordonate polare

$$I^2(\beta)=\int_{-\infty}^{\infty}e^{-\beta x^2}dx\int_{-\infty}^{\infty}e^{-\beta y^2}dy=\int_{-\infty}^{\infty}e^{-\beta\left(x^2+y^2\right)}dxdy=\int_0^{\infty}\int_0^{2\pi}e^{-\beta r^2}rdrd\varphi=\frac{\pi}{\beta}.$$

Prin urmare, rezultă $A=\sqrt{\beta/\pi}$.

Distribuția normală mai este cunoscută sub numele de *distribuție Gaus*s, sau simplu *Gaussiană* după numele marelui matematician german Karl Friedrich Gauss (1777-1855).

Generalizăm această distribuție pentru o funcție a cărei medie nu este nulă

$$f(x)=\sqrt{\beta/\pi}e^{-\beta(x-a)^2} .$$

Să calculăm media $\overline{x}$

$$\mu=\overline{x}=\sqrt{\beta/\pi}\int_{-\infty}^{\infty}xe^{-\beta(x-a)^2}dx=a .$$

În vederea calculării dispersiei să calculăm şi

$$\overline{x^2}=\sqrt{\beta/\pi}\int_{-\infty}^{\infty}x^2e^{-\beta(x-\mu)^2}dx=\frac{1}{2\beta}+\mu^2 ,$$

de unde

$$\sigma^2=\overline{x^2}-\left(\overline{x}\right)^2=\frac{1}{2\beta} .$$

Deci, în funcție de medie şi dispersie gaussiana se poate scrie

$$g(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

Momente inițiale şi momente centrate

Se numeşte *moment de ordinul s* al unei variabile aleatoare X media

$$\overline{(x-a)^s}=M\left[(x-a)^s\right]=\begin{cases}\sum_k x_kP_k, X\to \text{discretã}\\ \int_{-\infty}^{\infty}(x-a)^sf(x)dx, X\to \text{continuã}\end{cases},$$

dacă $a=0$ obținem momentul *inițial* de ordinul s (pe care-l notăm μ_s), iar dacă $a=\mu$ obținem momentul *centrat* de ordinul s (notat δ_s). În aceste notații avem $\mu_1=\mu=\overline{x}$, şi $\delta_2=\sigma^2=\overline{(x-\mu)^2}$. Legăturile dintre ele se deduc imediat

$\delta_1=\overline{(x-\mu)}=0$,

$\delta_2=\sigma^2=\overline{x^2}-\left(\overline{x}\right)^2=\mu_2-\mu_1^2$,

$\delta_3=\overline{(x-\mu)^3}=\mu_3-3\mu_2\mu_1+3\mu_1\mu_1^2-\mu_1^3=\mu_3-3\mu_2\mu_1+2\mu_1^3$, etc.

În cazul distribuției gauss momentele centrate capătă expresia

$$\delta_s = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^s e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx,$$

în care efectuăm schimbarea de variabilă $x-\mu=\sqrt{2}\sigma t = t/h$. Obținem astfel

$$\delta_s = \frac{1}{\sqrt{\pi}h^s}\int_{-\infty}^{\infty} t^s e^{-t^2}dt,$$

care integrată o dată prin părți conduce la relația de recurență

$$\delta_s = (s-1)\sigma^2\delta_{s-2}.$$

Cum $\delta_0 = 1$ și $\delta_1 = 0$ rezultă toate momentele centrate impare nule $\delta_{2k-1}=0$ iar cele pare sunt

$$\delta_{2k} = \frac{(2k-1)!}{2^k k!}\sigma^{2k-2}.$$

Coeficientul de asimetrie și de aplatizare. Prin definiție *coeficientul de asimetrie* (pe scurt, asimetria) al unei distribuții oarecare este definit ca

$$a = \frac{\delta_3}{\sigma^3},$$

iar coeficientul de aplatizare (excesul)

$$e = \frac{\delta_4}{\sigma^4} - 3.$$

Din relația de recurență specifică distribuției gaussiene rezultă că ambii coeficienți sunt nuli pentru o variabilă aleatoare cu distribuție normală.

Generarea numerelor aleatoare în calculatoare numerice

Cea mai simplă distribuție care se poate implementa în diferite limbaje de programare este *distribuția uniformă* (uniform deviate) în intervalul [0,1) a cărei funcție de distribuție este

$$U(x) = \begin{cases} 1, \text{pentru} & x\in[0,1) \\ 0, \text{in rest} & \end{cases}.$$

Pornind cu un număr întreg, precizat într-un mod oarecare înainte de fiecare apel (numit în literatura de limbă engleză *seed* - sămânță), rutina care generează numere aleatoare (random number generator) utilizează relația liniară de generare

$$RND = (aS+b) \bmod c,$$

unde *a*, *b* și *c* sunt constante întregi alese convenabil. Numărul întreg obținut astfel este considerat mantisa unui număr real în *reprezentarea virgulă mobilă*. Se obține astfel un număr real în intervalul [0,1). Modul cel mai uzual de a alege "sămânța" este utilizarea constantei întregi care memorează timpul curent înregistrat de ceasului calculatorului. Acest mod de generare nu produce, din punct de vedere riguros matematic, secvenţe de numere absolut aleatoare. Secvenţa de numere caracterizează o variabilă *pseudoaleatoare*. Spectrul acestei variabile este discret şi depinde atât de numărul de biţi alocaţi pentru un număr real în virgulă mobilă în calculator, precum şi de (foarte important) alegerea particulară a numerelor *a*, *b* şi *c*. Există alegeri defectuase celebre pentru cele trei constante. În principiul pentru numărul c trebuie ales un număr prim cât mai mare. Cea mai răspândită alegere, considerată drept un *generator standard minimal* este cea propusă de Park şi Miller [PAR 88]

$$a = 7^5 = 16807,\ b = 0 \text{ şi } c = 2^{31} - 1 = 2147483647$$

În unele variante de implementare a rutinei RND se utilizează o relaţia de recurenţă

$$RND = (aOLDRND + b) \bmod c,$$

iar utilizatorul precizează el care este numărul iniţial.