Gheorghe M.Panaitescu

MODELAREA SI SIMULAREA SISTEMELOR DE PRODUCTIE

Note de curs

Universitatea "Petrol-Gaze" Ploiesti
2006

P R E F A Ț Ă

Lucrarea prezentă este suportul de curs al disciplinei **Modelarea si simularea sistemelor de productie** si este destinată studentilor de la specializarea Inginerie economică în domeniul mecanic (IEDM), cursuri de zi dar si celor care urmează forma de învătământ la distantă.

Există si o versiune specială a acestor **Note de curs** dedicată învătământului la distantă, versiune care se aliniază instructiunilor de redactare elaborate de CNEAA.

Versiunea curentă a acestui curs cumulează o experientă dinamică care se măsoară în peste 10 ani de predare. An de an cursul a suferit modificări, de aceea a fost mentinut în formă electronică. Modificările si îmbunătătirile cele mai recente constau în suplimentarea cu un număr de exemple de modele si de calcule de simulare a unor sisteme de productie sau a unor părti ale lor.

Acest suport de curs este acompaniat de un volum separat de **Aplicatii si teste**, care contine un număr de teme si probleme propuse a fi solutionate în orele de aplicatii sau la pregătirea individuală precum si câteva teste de autoevaluare.

Pentru efectuarea calculelor necesare rezolvării unor aplicatii noi sau pentru verificarea exemplelor din corpul cursului, uneori este suficient un calculator de buzunar. Pentru calculele mai complexe autorul pune la dispozitia studentilor câteva programe de calcul (unele realizate chiar de autor) usor de instalat si de operat pe orice calculator personal. Cititorul găseste o prezentare a acestor programe în anexa volumului de **Aplicatii si teste**.

C U P R I N S

INTRODUCERE

Sistemele economice, la scară mică sau la scară mare au propriile lor legi cantitative care într-o economie normală, de piată nu pot fi ignorate si ar fi irational si contraproductiv să fie ignorate.
Istoric, legile acestea, mai simple sau mai complicate au fost sesizate de timpuriu si au fost mai întâi obiectul observării experimentale în scopul acumulării unor reguli empirice de a produce eficient, de a cheltui rezonabil si de a avea un profit bun în urma uneia sau alteia din activitătile de producere de bunuri destinate pietii sau de servicii.
Mai apoi, în ultimii vreo 75 de ani, inginerii si economistii, pe baza unei matematici particulare si sustinuti uneori nemijlocit de matematicieni au trecut la studii sistematice asupra comportării sistemelor economice, studii orientate în mare măsură spre crearea unor modele matematice care să cuprindă fenomenele specifice si să permită anticiparea, eventual îmbunătătirea performantelor economice.
O parte din preocupările acestea s-au desfăsurat în anii premergători si pe durata ultimului război mondial, sub forma unor studii asupra eficacitătii actiunilor de luptă. Ulterior, dar nu foarte târziu, s-a observat că multe din rezultatele de utilitate militară obtinute în acea perioadă sunt utilizabile si în domeniul economic. De aceea multe subiecte din cele regăsite în literatura din domeniu – la care cursul de fată încearcă a face o referire concisă si din care realizează si o sinteză fatalmente partială – se înscriu sub genericul cunoscut ca *Cercetarea operatională*, cuvântul al doilea al sintagmei provenind chiar de la *operatii* (militare) dar fiind nu mai putin adecvat unor operatii de natură economică.
Modelele prezentate în acest volum sunt în parte de tip determinist, în parte de tip stochastic, uneori un acelasi model continând în proportii variate atât aspecte deterministe cât si aspecte aleatoare, stochastice.
Modele de tip determinist sunt fără îndoială utile si de aceea primul capitol al lucrării are ca subiect central *modelele liniare de tip determinist* si mijlocul de tratare larg utilizat, metoda *programării liniare*.
Modelele deterministe sau părtile deterministe ale modelelor hibride determinist-stochastice sunt mai degrabă simplificări ale realitătii. Nu pot fi ignorate aspecte care în viata reală sunt fluctuante, sunt înafara exprimării prin numere exacte, sunt observabile dar trebuie tratate având în vedere mai multe valori pe care le pot lua aparent la întâmplare, incontrolabil. De aceea, în lucrare, urmează un capitol consistent de *probabilităti* si de *statisitică*

matematică util în întelegerea capitolelor următoare prin notiunile pe care le introduce sau le reaminteste.
O serie de realităti ale sistemelor economice tin seama de "istorie": ce se întâmplă la un moment dat depinde într-o manieră sau alta de ceea ce s-a întâmplat în perioada sau perioadele premergătoare. *Procesele Markov* sunt modele aproape ideale pentru aceste sisteme economice. Un capitol este consacrat modelării proceselor economice ca procese Markov.
Grafurile sunt modele de mare utilitate în modelarea sistemelor de productie. Problemele de conducere eficientă a executării unor lucrări de amploare, problemele de administrare judicioasă a retelelor de transport sunt mai usor de înteles si de solutionat dacă se utilizează teoria grafurilor. Un capitol este dedicat aplicatiilor acestor obiecte matematice, grafuri si retele, în modelarea sistemelor de productie.
Procesele de productie necesită uneori parcurgerea unui proces decizional preliminar sau de etapă. Un capitol al lucrării contine câtiva *algoritmi de elaborare a deciziilor*. O aplicatie în domeniu o are categoria specială de grafuri numiti arbori, sub forma arborilor decizionali.
În foarte multe sisteme economice relatiile sunt de natura client-server. Clientii solicitatori de un anumit serviciu sau de o sumă de servicii, persoane dar si piese, subansamble, echipamente întregi se pot afla în situatia de a astepta obtinerea serviciului necesar. *Sistemele cu asteptare* sunt obiectul unui capitol tocmai datorită importantei echilibrării costurilor asteptării cu costurile servirii.
Între modalitătile de simulare a sistemelor de productie, simularea aleatoare este de foarte mare utilitate în întelegerea sistemelor care includ aspecte stochastice. Un capitol al lucrării este dedicat *simulării* Monte Carlo.
Programarea judicioasă a productiei necesită si unele proiectii în viitor ale conditiilor de desfăsurare a procesului productiv. Sunt necesare, asadar, *prognoze economice*. Prognozele sunt tratate într-un capitol aparte.
În ultimii ani se vorbeste foarte frecvent de sistemele de fabricatie flexibile. Flexibilitatea se referă atât la alternativele multiple de implementare posibile ale tehnologiei unui produs cât si la rapiditatea trecerii de la un tip de produs la altul într-un sistem cu aceleasi dotări materiale. *Retelele Petri* sunt un model de foarte mare utilitate în proiectarea si operarea controlată a sistemelor flexibile de fabricatie. Un capitol întreg face referiri suficient de ample la acest gen de modele.
Nu s-au evitat nici metodele mai noi de modelare si simulare a sistemelor productive. Un ultim capitol al acestei lucrări contine referiri concise la modelele de tipul *retelelor neuronale* si la mijloacele de rezolvare a unor probleme de modelare si conducere a sistemelor de productie pe bază de model.
Lucrarea contine un număr de exemple de modele si exemple de calcule pe baza acestor modele.
Aproape fiecare capitol are în partea finală un număr de probleme propuse studentului spre rezolvare. Unele dintre acestea se pot rezolva "în vârful creionului", adică nu este nevoie de mai mult decât un calculator de buzunar, uneori nici măcar de atât. Altele necesită programe de calcul mai rafinate si, de

aceea, autorul pune la dispozitia studentilor, concomitent cu textul lucrării de fată, câteva programe de calcul cu care se pot rezolva problemele din această categorie (SP01, SP04, SP07).
Fiecărui capitol i s-au atsat *exercitii de autoevaluare* sub forma unor întrebări însotite de mai multe răspunsuri din care trebuie ales cel corect. La sfârsitul lucrării sunt date spre verificare răspunsurile corecte ale acestor exercitii.
În completare, alte exemple numerice sunt propuse pentru orele de aplicatii. **Ghidul de lucrări** la disciplina **Modelarea si simularea sistemelor de productie** reprezintă o sursă importantă de probleme de solutionat. Încă mai bogate sunt sursele bibliografice indicate.
În siteza celor scrise mai sus, se poate afirma că lucrarea prezentă este un suport utilizabil în cunoasterea problemelor principale ale modelării si simulării sistemelor de productie si un start consistent în aprofundarea ulterioară a cunostintelor în domeniu.

MODELE LINIARE DE TIP DETERMINIST

Sistemele de productie sunt descrise adesea de sisteme de relatii liniare. Liniaritatea se referă la calitatea relatiilor constitutive de a contine variabilele modelului, adesea numite variabile de decizie, exclusiv la puterea întâia.
Există, desigur, si sisteme de relatii modelatoare care nu sunt liniare în variabilele lor. Ele se pot în general liniariza prin formula lui Taylor cunoscută de la matematici, limitată la termenii de gradul întâi. Singura conditie care trebuie îndeplinită de fiecare din functiile componente ale modelului este aceea de a fi derivabile până la ordinul al doilea inclusiv. Atunci formula

$$F(x_1, x_2, \ldots, x_n) = F(x_{10}, x_{20}, \ldots, x_{n0}) + \sum_{i=1}^{n} \frac{\partial F}{\partial x_i}(x_i - x_{i0}) + R_2$$

cu derivatele partiale evaluate în punctul $(x_{10}, x_{20}, \ldots, x_{n0})$ liniarizează functia pe un domeniu în jurul acestui punct, domeniu în care restul formulei, R_2 este contextual tolerabil de mic.
Modelele liniare prin natura lor sau cele obtinute prin liniarizarea unor ecuatii neliniare fac adesea obiectul unor preocupări de optimizare. De mare utilitate în optimizarea sistemelor economice este metoda programării liniare prezentată în continuare. Se-ntelege aproape de la sine că rezultatele optimizării pe modele care contin ecuatii obtinute prin liniarizare trebuie tratate cu prudentă, în sensul de a vedea dacă este asigurată compatibilitatea cu ecuatiile originare neliniare.
Problema optimizării se pune si pentru modele neliniare. Aceasta va fi adusă în discutie ceva mai departe.

Programarea liniară

În matematică, programarea liniară este o metodă de localizare a posibilelor extreme ale functiilor liniare într-un spatiu delimitat de suprafete exprimate la rândul lor prin ecuatii liniare. Au fost create metode robuste de a rezolva problemele de acest gen si există implementări bine puse la punct ale algoritmilor de calcul corespunzători. Inevitabil, trebuie amintit aici algoritmul *simplex*.
Cum s-a spus deja, multe dintre modelele sistemelor de productie sunt prin natura lor liniare, iar multe altele pot fi aproximate pe zone mai largi sau mai restrânse ale spatiului variabilelor care le descriu, prin relatii liniare. Este de asteptat, asadar, ca multe probleme de optimizare legate de activitatea economică să fie tratate cu mijloacele puse la dispozitie cu generozitate de programarea liniară.

Formularea generală a problemelor de programare liniară are în vedere o *functie obiectiv*

$$c_0 + c_1x_1 + c_2x_2 + \ldots + c_nx_n$$

care trebuie minimizată (sau alteori maximizată) prin stabilirea unor valori adecvate ale *variabilelor de decizie* $x_1, x_2, \ldots, x_n$ care să respecte în acelasi timp un set de *restrictii* de forma

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1$$
$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m$$

în număr de *m* si, foarte frecvent, *conditia de nenegativitate* a valorilor varibilelor de decizie: $x_1 \ge 0, x_2 \ge 0, \ldots, x_n \ge 0$.

Desigur, inegalitătile pot fi uneori egalităti, alteori pot fi inegalităti în sensul celălalt, mai mare sau egal, dar prin înmultirea cu –1, operatie permisă, se pot aduce la forma de mai sus.

Este de retinut un fapt: utilizarea programării liniare este în mare măsură o problemă cu multe aspecte de ordin practic. De aceea, fără altă amânare, sunt prezentate în continuare câteva exemple de probleme ale căror solutii se pot obtine prin programare liniară.

Exemplul 1. **O problemă de planificare a productiei**

O societate comercială realizează un produs în patru variante si în partea finală a procesului de productie sunt necesare operatii de asamblare, de vopsire-lustruire si de împachetare. Pentru fiecare variantă a produsului, operatiile mentionate consumă durate diferite de la caz la caz si profitul specific este diferit asa cum se poate vedea în tabelul care urmează.

Varianta	Operatii si durate (min)			Profit (u.m.)
	Asamblare	Vopsire-lustruire	Împachetare	
1	2	3	2	1,50
2	4	2	3	2,50
3	3	3	2	3,00
4	7	4	5	4,50

- Dată fiind starea curentă a fortei de muncă angajate, se estimează că anual sunt disponibile 100.000 de minute pentru asamblare, 50.000 de minute pentru operatia de vopsire-lustruire si 60.000 de minute pentru împachetare. Câte unităti fizice trebuie produse anual din fiecare variantă si care ar fi profitul realizat?
- Admitând că există posibilitatea de a decide liber asupra timpului dedicat celor trei operatii, în interiorul timpului alocabil total de 210.000 (= 100.000 + 50.000 + 60.000) minute, câte unităti fizice trebuie produse anual din fiecare variantă si care ar fi profitul realizat?

Desigur, producătorul este interesat ca profitul să fie cât mai mare.
Problema reformulată în relatii. Fie x_i (i = 1, 2, 3, 4) productia anuală în unităti fizice din fiecare variantă a produsului, fie T_a, T_{vp}, T_p numărul de minute utilizate anual pentru asamblare, vopsire-lustruire, respectiv împachetare. Toate variabilele definite aici iau obligatoriu valori pozitive sau nule $x_i \geq 0$ (i = 1, 2, 3, 4), $T_a \geq 0$, $T_{vp} \geq 0$, $T_p \geq 0$ si între ele au loc următoarele relatii:
$T_a = 2x_1 + 4x_2 + 3x_3 + 7x_4$ (timpul total consumat anual cu asamblarea)
$T_{vp} = 3x_1 + 2x_2 + 3x_3 + 4x_4$ (timpul total consumat anual cu vopsirea-lustruirea)
$T_p = 2x_1 + 3x_2 + 2x_3 + 5x_4$ (timpul total consumat anual cu împachetarea)
În primul caz timpul disponibil specific pentru cele trei operatii este limitat si de aici rezultă restrictiile

$$T_a = 2x_1 + 4x_2 + 3x_3 + 7x_4 \leq 100.000$$
$$T_{vp} = 3x_1 + 2x_2 + 3x_3 + 4x_4 \leq 50.000$$
$$T_p = 2x_1 + 3x_2 + 2x_3 + 5x_4 \leq 60.000$$

În cazul al doilea suma $T_a + T_{vp} + T_p$ este limitată si din această limitare rezultă restrictia

$$T_a + T_{vp} + T_p = 7x_1 + 9x_2 + 8x_3 + 16x_4 \leq 210.000$$

Profitul este functia obiectiv si depinde de valorile variabilelor de decizie x_i (i = 1, 2, 3, 4) conform relatiei

$$P = 1{,}5x_1 + 2{,}5x_2 + 3{,}0x_3 + 4{,}5x_4$$

Se poate observa usor similitudinea enuntului prezentat în acest paragraf cu formularea generală a problemei prezentată în partea introductivă a acestei sectiuni.
Exemplul 2. **O problemă de planificare financiară**
O bancă acordă clientilor săi patru tipuri de credite care aduc anual următoarele dobâzi:

- Credite ipotecare initiale: 14%
- Credite ipotecare secundare: 20%
- Credite pentru îmbunătătiri ale locuintelor: 20%
- Credite pentru acoperirea depăsirilor de disponibil în cont: 10%

Banca are o capacitate de creditare estimată la 250.000.000 u.m. si îsi propune să facă fată următoarelor elemente de politică vis-à-vis de clienti:

- Creditele ipotecare initiale trebuie să fie de cel putin 55% din totalul creditelor ipotecare acordate si cel putin 25% din totalul creditelor acordate
- Creditele ipotecare secundare nu pot depăsi 25% din totalul creditelor acordate
- Pentru a evita disconfortul clientilor si/sau introducerea unor taxe surpriză pe parcursul derulării creditelor, dobânda medie pentru toate creditele acordate trebuie să nu depăsească 15%

Cu toate că aceste măsuri limitează profitul pe care banca l-ar putea avea, ele au menirea de a proteja banca fată de riscurile excesive pe care un aspect particular le-ar putea crea. De aceea, interesul băncii este să maximizeze veniturile din dobânzile pretinse la credite, în conditiile respectării politicii de creditare enuntate mai sus.

Si în cazul acestei probleme trebuie definite *variabilele de decizie*, *restrictiile* si *functia obiectiv*.

Variabilele x_i (i = 1, 2, 3, 4) sunt sumele pe care banca le acordă pe cele patru categorii de credite, în ordinea din enunt. Valorile acestora nu pot fi negative, asadar $x_i \geq 0$ (i = 1, 2, 3, 4).

Restrictiile vin din:

- Suma totală a creditelor

$$x_1 + x_2 + x_3 + x_4 \leq 250$$

- Conditia primă din politica băncii

$$x_1 \geq 0,55(x_1 + x_2)$$
$$x_1 \geq 0,25(x_1 + x_2 + x_3 + x_4)$$

- Conditia a doua din politica băncii

$$x_2 \leq 0,25(x_1 + x_2 + x_3 + x_4)$$

- Conditia a treia din politica băncii

$$0,14x_1 + 0,20x_2 + 0,20x_3 + 0,10x_4 \leq 0,15(x_1 + x_2 + x_3 + x_4)$$

Relatiile restrictive care cuprind politica băncii nu sunt, după cum se vede, în forma standard. Înainte de rezolvare efectivă a problemei, acestea trebuie să fie aduse la forma standard. La acest moment forma standard ar afecta semnificatia concretă a acestor relatii. Este de observat însă forma lor liniară.

Functia obiectiv exprimă venitul total din dobânzi

$$0,14x_1 + 0,20x_2 + 0,20x_3 + 0,10x_4$$

care trebuie maximizat.

O solutie a problemei este x_1 = 208,33, x_2 = 41,67 si $x_3 = x_4 = 0$. Rezultatul acesta este obtinut pe calculator prin executia unui anumit program. Despre solutie se poate spune mai întâi că satisface toate restrictiile problemei formulate mai devreme. Dar solutia aceasta nu este unică. O a doua solutie, x_1= 62,50, x_2 = 0, x_3 = 100, x_4 = 87,50 satisface si ea toate restrictiile si conduce la aceeasi valoare maximă (de 37,5) ca si solutia primă, după cum se poate verifica. Este de retinut din acest exemplu că o problemă de programare liniară poate admite uneori mai multe solutii cu valori ale functiei obiectiv egale.

Exemplul 3. **O problemă de realizare a unor amestecuri**

Un producător de furaje are de rezolvat probleme de genul care urmează.

Pentru vacile de lapte trebuie să sintetizeze un furaj prin amestecarea a două ingrediente active într-un furaj standard de bază, care constituie componenta principală. Un kilogram de furaj-amestec trebuie să contină cantităti minime din patru principii nutritive, după cum urmează:

Principiul nutritiv	A	B	C	E
g/kg	90	50	20	2

Componentele active contin principiile nutritive în proportii date si au anumite costuri. Aceste valori sunt prezentate în tabelul următor:

	A	B	C	D	Cost/kg
Ingredientul 1 (g/kg)	100	80	40	10	40
Ingredientul 2 (g/kg)	200	150	20	–	60

Care ar trebui să fie cantitătile componentelor active si ale adaosului de completare într-un kilogram de furaj?
Aici, *variabilele* sunt în număr de trei:
x_1 = cantitatea (kg) din ingredientul 1 într-un kilogram de furaj
x_2 = cantitatea (kg) din ingredientul 2 într-un kilogram de furaj
x_3 = cantitatea (kg) de furaj de bază într-un kilogram de furaj
Aceste cantităti nu pot fi negative: $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$ si formează un gen de retetă a preparării furajului conform normelor enuntate mai sus.
Restrictiile au sursa în continutul minim recomandat al furajului în fiecare din principiile nutritive A, B, C si D

$100x_1 + 200x_2 \geq 90$ (principiul A)
$80x_1 + 150x_2 \geq 50$ (principiul B)
$40x_1 + 20x_2 \geq 20$ (principiul C)
$10x_1 \geq 2$ (principiul D)

si în conditia de normalizare

$x_1 + x_2 + x_3 = 1$

Functia obiectiv este dată de costul retetei

$40x_1 + 60x_2$

care trebuie minimizat.
Solutia problemei este x_1 = 0,3667, x_2 = 0,2667 si x_3 = 0,3667 exprimată cu patru cifre după virgulă.
Problema are posibile extinderi prin cresterea numărului de principii nutritive, prin cresterea numărului de ingrediente amestecate, prin modificări ale limitelor de continut în principii active, prin tratarea variatiilor de costuri, oricând posibile, prin tratarea unor dificultăti în aprovizionare, prin includerea costului furajului de bază.
Problemele legate de dieta umană pot fi tratate într-o manieră similară.
Exemplul 4. **O altă problemă de planificare industrială**
În conditii normale, o fabrică produce până la 100 de unităti fizice dintr-un anumit produs în patru intervale de timp succesive (de pildă trimestre) la costuri care se modifică de la interval la interval conform tabelului de mai jos.
Prin utilizarea unor ore suplimentare de lucru, se poate obtine un plus de productie. Cantitătile maxime si costurile sunt prezentate tot în tabelul care urmează, alături de prognoza asupra cererii în fiecare din cele patru intervale.
Este posibil a se depozita până la 70 de unităti fizice de la o perioadă la alta la un cost de 1,5 u.m. (unităti monetare) pe perioadă (Numărul 1,5 u.m./perioadă este cunoscut sub numele de cheltuieli de stocare).

Intervalul	Cererea (u.f.)	Costuri normale (u.m./u.f.)	Productie peste capacitate (u.f.)	Costuri cu productia suplimetară (u.m./u.f.)
1	130	6	60	9
2	80	4	65	6
3	125	8	70	10
4	195	9	60	11

Se cere determinarea productiei si planului de stocare care se potrivesc cererii în cele patru perioade în conditii de cost total minim. La începutul primei perioade există un stoc initial de 15 unităti fizice (u.f.).
Formularea acestei probleme ca o problemă de programare liniară parcurge etapele următoare:
Variabile.
Numărul de unităti produse prin efort normal în cele patru perioade, x_t (t = 1, 2, 3, 4) , $x_t \geq 0$
Numărul de unităti produse prin efort suplimentar în cele patru perioade, y_t (t = 1, 2, 3, 4) , $y_t \geq 0$
Numărul de unităti în stoc la sfârsitul fiecărei perioade, I_t (t = 1, 2, 3, 4)
Restrictii.
Din limitările de productie: $x_t \leq 100$ (t = 1, 2, 3, 4), $y_1 \leq 60$, $y_2 \leq 65$, $y_3 \leq 70$, $y_4 \leq 60$.
Din limitarea sapatiului de stocare: $I_t \leq 70$ (t = 1, 2, 3, 4)
Din conditia de continuitate: stocul de închidere = stocul de deschidere + productia – cererea
Se presupune că stocul de deschidere a perioadei t este egal cu stocul de închidere a perioadei premergătoare si că productia în intervalul t este de natură să acopere cererea din perioada t. Se scriu relatiile:
$I_1 = I_0 + (x_1 + y_1) - 130$
$I_2 = I_1 + (x_2 + y_2) - 80$
$I_3 = I_2 + (x_3 + y_3) - 125$
$I_4 = I_3 + (x_4 + y_4) - 195$
cu $I_0 = 15$.
Ecuatiile de continuitate a stocului srise mai sus sunt tipice pentru problemele de planificare care se referă la mai mult de un interval de timp. Variabilele de inventar I_t si ecuatiile de continuitate a stocului pun în legătură intervalele avute în vedere si reprezintă o evidentă fizică a stocului.
Cererea trebuie totdeauna satisfăcută. Nu este permisă lipsa de stoc. Asta se poate spune si altfel, echivalent: stocul de deschidere pentru perioada t plus productia din acea perioadă trebuie să fie cel putin cât cererea în perioada t, adică:
$I_0 + (x_1 + y_1) \geq 130$
$I_1 + (x_2 + y_2) \geq 80$

$I_2 + (x_3 + y_3) \geq 125$
$I_3 + (x_4 + y_4) \geq 195$
Aceste ecuatii pot fi văzute si altfel. Luând în considerare ecuatiile de continuitate a stocului, ecuatiile de mai sus care asigură satisfacerea întotdeauna a cererii se mai pot scrie ca:
$I_1 \geq 0$
$I_2 \geq 0$
$I_3 \geq 0$
$I_4 \geq 0$
Functia obiectiv. Costul are trei componente: costul productiei realizate în conditii normale, costul productiei realizate prin muncă suplimentară si costul reportării de stocuri de la o perioadă la următoarea:

$$(6x_1 + 4x_2 + 8x_3 + 9x_4) + (8y_1 + 6y_2 + 10y_3 + 11y_4) +$$
$$+ (1{,}5I_0 + 1{,}5I_1 + 1{,}5I_2 + 1{,}5I_3 + 1{,}5I_4)$$

Costul acesta trebuie minimizat.
Rezultatele ar trebui să fie numere întregi, dar pot fi si numere reale. Dacă numărul de unităti cerut în fiecare perioadă este mare atunci fractiile s-ar putea să nu deranjeze prea mult.
Solutia acestei probleme de programare liniară este

$$x_1 = x_2 = x_3 = x_4 = 100$$
$$y_1 = 15, y_2 = 50, y_3 = 0 \text{ si } y_4 = 50$$
$$I_0 = 15, I_1 = 0, I_2 = 70, I_3 = 45 \text{ si } I_4 = 0.$$

Valoarea minimă a functiei obiectiv este 3865.
Exemplele enuntate mai sus sunt propuse cititorului spre rezolvare în orele de aplicatii sau ca temă de casă pe calculatorul personal. Cu acest prilej se pot verifica si solutiile date mai devreme, dar si variatia rezultatelor în urma unor modificări ale coditiilor din enunturi. Alte câteva probleme sunt propuse în continuare.

ELEMENTE DE TEORIA PROBABILITĂTILOR SI DE STATISTICĂ MATEMATICĂ

Aproape toate fenomenele economice sunt însotite de aspecte aleatoare. Asadar, în particular, în modelele sistemelor de productie trebuie incluse elemente care să facă posibilă interpretarea corectă a efectelor pe care întâmplarea le poate avea asupra functionării unui sistem economic. De aceea, această a doua parte a cursului de *Modelarea si simularea sistemelor de productie* contine un minim de referiri la teoria probabilitătilor si la statistica matematică.

Spatiul evenimentelor

Un experiment oarecare, provocat sau spontan poate avea rezultate diverse. Aceste rezultate sunt denumite mai departe *evenimente*. Astfel, rostogolirea unui zar corect pe o suprafată plană orizontală (exemplu foarte banal dar des utilizat de matematicieni) poate avea ca rezultat aparitia pe fata sa de deasupra a, să spunem, cinci puncte. S-a produs asadar evenimentul aparitiei pe feta de deasupra a cinci puncte. Tot asa, conform definitiei de mai sus, extragerea valetului de cupă dintr-un pachet de cărti de joc bine amestecat este un eveniment.

Dac**ă** E este multimea tuturor evenimentelor posibile relativ la un experiment, aceast**ă** multime poate fi numită, cum adesea se întâmpl**ă,** *spatiul evenimentelor*. Evenimentele unui astfel de spatiu se pot găsi în anumite relatii unul cu altul, cu evenimentele acelui spatiu se pot face unele operatii.

O relatie importantă între evenimente este *implicatia*. Implicatia se notează $A \subset B$ si se citeste *evenimentul A implică evenimentul B*, ceea ce înseamnă că producerea evenimentului A conduce automat, implicit la producerea evenimentului B; implicatia reciprocă, $A \subset B$ si $B \subset A$ este un mod de a exprima egalitatea (echivalenta) a două evenimente.

Operatiile cu evenimente sunt *unare*, cu un singur eveniment ca operand, sau *binare*, cu două evenimente ca operanzi.

Operatia de luare a *complementarului* sau, ceea ce este totuna, a *contrarului* unui eveniment este unară, operează cu un singur eveniment.

Reuniunea si *intersectia* de evenimente sunt operatii binare, implică două evenimente.

Luarea complementarului sau a contrarului unui eveniment constă în luarea în considerare a acelui eveniment care se produce atunci când nu se produce evenimentul al cărui contrar se caută. Într-un exemplu foarte simplu, aruncarea unei monede cu cădere pe o suprafată plană orizontală poate avea ca rezultat afisarea deasupra fie a unei fete, fie a celeilalte. Fiecare din cele două evenimente generate de acest experiment simplu este contrarul celuilalt. Dacă evenimentul asupra căruia se operează este A atunci evenimentul contrar este notat cu $\overline{A}$. De ce *contrar* si/sau *complementar* se va explica mai în detaliu după definirea celor două operatii binare anuntate.

Reuniunea a două evenimente se notează $A \cup B$ si este evenimentul care constă în producerea a cel putin unuia din cele două evenimente, adică sau a unui sau a celuilalt sau a ambelor evenimente.

Intersectia a două evenimente se notează $A \cap B$ si este evenimentul constând în producerea ambelor evenimente, adică atât a unui eveniment cât si a celuilalt.

Există două evenimente speciale care se includ în multimea E. Unul este evenimentul imposibil, notat cu $\varnothing$, evenimentul care nu se produce niciodată. Celălalt este evenimentul sigur, notat cu E, evenimentul care se produce de fiecare dată.

O relatie de forma $A \cap B = \varnothing$ exprimă incompatibilitatea reciprocă a celor două evenimente A si B, în alte cuvinte producerea unuia exclude producerea celuilalt.

Acum se poate formula mai precis raportul între un eveniment si contrarul lui: $A \cap \overline{A} = \varnothing$, $A \cup \overline{A} = E$. Într-o lectură în cuvinte a acestor relatii: un eveniment este incompatibil cu contrarul său, producerea unui eveniment *sau* a contrarului său este sigură. Este acum momentul să se aducă precizarea că contrarul contrarului unui eveniment este acel eveniment. Simbolic, aceasta se scrie $\overline{\overline{A}} = A$.

Multimea E este partial ordonată, relatia de ordine este implicatia.

Multimea E împreună cu operatiile de luare a complementarului unui eveniment, de reuniune si de intersectie a evenimentelor se organizează ca o algebră booleană.

Între evenimentele dintr-o multime E se disting atomi (sau evenimente elementare) si evenimente compuse. De pildă, prin aruncarea zarului se pot produce între altele evenimentele A_2 si A_5 care constau în aparitia pe fata de deasupra a numărului de puncte trecut ca indice. Ambele sunt atomi sau evenimente elementare în sensul că nu sunt alte evenimente încă mai simple decât ele. Reuniunea $A_2 \cup A_5$ este însă un eveniment compus.

Fie acum Ω multimea tuturor evenimentelor elementare dintr-o multime finită E de evenimente. Evident $\Omega \neq \varnothing$. O submultime de părti ale lui Ω, $K \subset P(\Omega)$ se organizează ca un corp dacă

$$A \in K \Rightarrow \overline{A} \in K$$
$$A, B \in K \Rightarrow A \cup B \in K$$
$$A, B \in K \Rightarrow A \cap B \in K$$

În aceste coditii perechea (Ω, K) este un corp de evenimente si este un σ–corp sau corp borelian de evenimente dacă orice reuniune sau intersectie finită sau infinită de evenimente din K apartine multimii K.

Într-un spatiu E complet si atomic, orice eveniment $A \neq \varnothing$ se poate scrie ca o reuninune de elemente din Ω

$$A = \bigcup_{\omega_i \in \Omega} \omega_i$$

Se numeste *partitie* a unui eveniment $A \in K$ o multime de evenimente $A_i \in K, (i = 1,2,...,n)$, care sunt mutual incompatibile, adică $A_i \cap A_j = \varnothing$ pentru orice pereche $A_i, A_j \in K$ cu $i \neq j$, astfel încât

$$\bigcup_{i=1}^{n} A_i = A$$

Dacă $A = \Omega$ atunci evenimentele $A_i \in K, (i = 1,2,...,n)$ alcătuiesc un *sistem complet de evenimente*.

Probabilităti, probabilităti conditionate

Pe multimea evenimentelor din K se defineste o functie reală P, numită *probabilitate*, care are proprietătile:

1. $P(A) \geq 0$, $\forall A \in K$
2. $P(\Omega) = 1$
3. $P\left(\bigcup A_i\right) = \sum P(A_i),\ A_i \in K,\ A_i \cap A_j = \varnothing, i \neq j$

Dacă ultima proprietate are loc si pentru reuniuni numerabile atunci probabilitatea P se numeste complet aditivă (sau σ–aditivă) pe corpul (borelian) de evenimente (Ω, K).
Tripletul (Ω, K, P) se numeste câmp (borelian) de probabilitate. Dacă Ω este o multime finită atunci (Ω, K, P) este un câmp de probabilitate discret.
Din proprietătile de mai sus derivă alte câteva proprietăti importante ale probabilitătii P. Astfel

4. $P(\varnothing) = 0$
5. $P(A) = 1 - P(\overline{A})$
6. $P(A - B) = P(A) - P(A \cap B)$
7. $0 \leq P(A) \leq 1$
8. $P(A \Delta B) = P(A) + P(B) - 2P(A \cap B)$
9. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

unde $A - B = A \cap \overline{B}$ si $A \Delta B = (A - B) \cup (B - A)$ sunt diferenta, respectiv diferenta simetrică a două evenimente. O extindere a relatiei ultime la reuniunea a n evenimente este

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{j=1}^{n} (-1)^{j+1} S_j \quad \text{cu} \quad S_j = \sum_{i_1, i_2, ..., i_j \leq n} P(A_{i_1} \cap ... \cap A_{i_j}) \ \ j \leq n .$$

Dacă $F = \{A_i\}_{i \in I}$ este o familie numerabilă de evenimente mutual incompatibile, atunci $P\left(\bigcap_{i \in I} A_i\right) = 0$.

Dacă familia $F = \{A_i\}_{i \in I}$ este si exhaustivă, adică se constituie ca un sistem complet de evenimente, atunci $P\left(\bigcup_{i \in I} A_i\right) = 1$.

Evenimentele se pot afla în relatie de conditionare reciprocă în sensul că un eveniment odată produs poate modifica probabilitatea de producere a altui eveniment. Relatia de bază pentru calculul probabilitătilor conditionate este

$$P_B(A) = P(A/B) = P(A \cap B)/P(B)$$

cu evenimentul B, cel care conditionează producerea evenimentului A, trecut ca indice sau pe pozitia a doua, după caracterul "/", în argumentul functiei probabilitate.

În general,

$$p(A/B) \neq P(A) \text{ si } P(B/A) \neq P(B)$$

ceea ce indică o dependentă, o conditionare între cele două evenimente. Dacă are loc egalitatea în ambele relatii atunci evenimentele nu se conditionează în nici un fel, sunt independente.

Dacă probabilitatea unei intersectii finite de evenimente este nenulă

$$P\left(\bigcap_{i=1}^{n} A_i\right) \neq 0$$

atunci probabilitatea respectivă se poate calcula cu formula

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P\left(A_n / \bigcap_{i=1}^{n-1} A_i\right) \ldots P(A_2/A_1)P(A_1)$$

care se demonstrează inductiv pornind de la relatia pentru două evenimente derivată din formula probabilitătii conditionate

$$P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$$

Dacă $F = \{A_i\}_{i=\overline{1,n}}$ este o partitie a câmpului Ω atunci probabilitatea unui eveniment oarecare se poate calcula cu relatia

$$P(A) = \sum_{i=1}^{n} P(A_i)P(A/A_i)$$

cunoscută sub numele de *formula probabilitătii totale*.

Mai este de retinut *formula lui Bayes*

$$P(A_i/A) = \frac{P(A_i)P(A/A_i)}{\sum_{i=1}^{n} P(A_i)P(A/A_i)}$$

care în aceleasi conditii, $F = \{A_i\}_{i=\overline{1,n}}$ o partitie a câmpului Ω, permite calculul probabilitătii fiecărui eveniment al partitiei conditionat de evenimentul $A \in K$, altfel oarecare.

Exemplu. În cazul zarului corect enuntat mai devreme, multimea Ω este alcătuită din evenimentele A_1, A_2, A_3, A_4, A_5, A_6. Multimea de părti ale lui Ω care se constituie în corp de evenimente este multimea tuturor reuniunilor de 2, de 3, de 4, de 5 sau de 6 evenimente la care se adaugă evenimentele atomice, elementare deja enumerate si evenimentul imposibil $\varnothing$.

Prin perceptie imediată se poate afirma că sansele de producere a celor sase evenimente sunt egale (sansa aceasta de producere a unui eveniment sau a altuia este măsurată de probabilitate). Se poate scrie, asadar

$$P(A_1) = P(A_2) = P(A_3) = P(A_4) = P(A_5) = P(A_6) = p$$

Evenimentul sigur Ω se poate scrie ca o reuniune

$$\Omega = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$$

si deoarece evenimentele din reuniune sunt două câte două mutual incompatibile (nu pot apărea deasupra două fete diferite deodată) conform proprietătii 3 se poate scrie

$$1 = P(\Omega) = P(A_1) + P(A_2) + P(A_3) + P(A_4) + P(A_5) + P(A_6) = 6p$$

adică $p = 1/6$. Acum se pot calcula probabilităti diverse.

a) Probabilitatea aparitiei unui număr par de puncte este

$$P(A_2 \cup A_4 \cup A_6) = P(A_2) + P(A_4) + P(A_6) = 3(1/6) = 1/2$$

ca probabilitate a unei reuniuni de evenimente două câte două reciproc incompatibile.

b) Probabilitatea evenimentului A_4 conditionată de evenimentul reuniune de la punctul precedent, $A = A_2 \cup A_4 \cup A_6$

$$P(A_4/A) = \frac{P(A_4 \cap A)}{P(A)} = \frac{P[A_4 \cap (A_2 \cup A_4 \cup A_6)]}{P(A)} = \frac{P(A_4)}{P(A)} = \frac{1/6}{1/2} = \frac{1}{3}$$

etc.

De retinut din acest exemplu o modalitate de a evalua probabilităti prin raportarea numărului de situatii favorabile unui eveniment la numărul total de situatii. De pildă, evenimentul A_3 din cele de mai sus se produce în proportia 1 caz favorabil din 6 posibile, adică 1/6.

La loteria “6 din 49”, se pot juca C_{49}^6 (combinări de 49 de numere luate câte 6) variante distincte. Sansa (probabilitatea) unei variante particulare de a fi câstigătoare a premiului cel mare este de $1/C_{49}^6$, o probabilitate foarte, foarte mică, desigur.

Sansa de a câtiga la categoria a II-a este întrucâtva mai mare. Un bilet câtigător poate contine una din cele $C_6^5 = 6$ combinatii de 5 numere din cele iesite din urnă la extragerea duminicală, la care se adaugă unul din celelalte 43 de numere din afara extragerii. Numărul de situatii favorabile câstigării unui premiu la categoria a II-a este, evident, 6x43 = 258 si probabilitatea este de $258/C_{49}^6$, încă destul de mică dar mai mare decât cea de la categoria I.

La jocul de table (foarte cunoscut în toată lumea – backgammon), evolutia disputei dintre jucători este hotărîtă pas cu pas prin aruncarea a două zaruri. Într-o anumită fază a jocului, unul dintre jucători are nevoie ca zarurile să producă o sumă a punctelor egală cu 5. Care este probabilitatea acestui eveniment? Numărul total de rezultate este 36: fiecare din cele sase fete ale unui zar poate apărea combinată cu oricare din cele sase fete ale celuilalt zar. Suma punctelor este 5 în următoarele 4 cazuri: (1, 4), (2, 3), (3, 2) si (4, 1). Prin raportarea numărului de cazuri favorabile (4) la numărul total de cazuri (36) se obtine răspunsul la întrebare: 4/36 = 1/9.

Variabile aleatoare

Formal, o *variabilă aleatoare* este o functie definită pe o multime atomică cu valori reale, $X: \Omega \to R$, care are proprietatea

$$\{X < x\} \Rightarrow \{\omega \in \Omega / X(\omega) < x\} \in K \text{ , } \forall x \in R$$

În termeni mai putini rigurosi din punct de vedere matematic, o variabilă aleatoare este o variabilă care ia valori la întâmplare dar în nici un caz haotic. Explicit sau tacit, în spatele manifestării varibilei aleatoare prin valori diverse se află un câmp de probabilitate (Ω, *K*, *P*) definit de multimea atomică Ω, de corpul de părti ale acesteia *K* si de probabilitatea *P*. Probabilitatea face ca unele valori pe care variabila aleatoare le poate lua să fie (eventual) mai probabile decât altele. Numărul de pucte afisate de un zar comun este o variabilă aleatoare. Cu fetele zarului, care pot fi de pildă colorate, nu neapărat "punctate", se pot asocia si alte numere printr-o functie *X* particulară. Functia *X* este o altă variabilă aleatoare definită pe câmpul (Ω, *K*, *P*).

O *variabilă aleatoare simplă* ia numai un număr finit de valori. De exemplu functia indicator a unui eveniment $A \in K$, care se poate produce sau nu

$$\chi_A = \begin{cases} 0 & \omega \notin A \\ 1 & \omega \in A \end{cases}$$

este o variabilă aleatoare simplă care ia numai două valori, 0 si 1. Variabilele aleatoare definite în relatie cu zarul sunt variabile aleatoare simple.

Dacă *X* este o variabilă aleatoare definită pe câmpul (Ω, *K*, *P*) atunci pentru oricare două valori $x_1, x_2 \in R,\ x_1 \le x_2$ toate intervalele finite sau infinite delimitate de cele două valori corespund unor evenimente din *K* si, prin generalizare, pentru orice multime *I* reuniune de intervale din *R*, se poate calcula $P_X(I) = P[X(\omega) \in I] = P[X^{-1}(I)]$.

$P_X(I)$ reprezintă *distributia de probabilitate* a variabilei aleatoare *X*. Se poate vorbi de P_X ca de o probabilitate definită pe câmpul (*R*, K_X) în care $K_X = \{I \subset R / X^{-1}(I) \in K\}$ este o multime de intervale ale multimii numerelor reale *R*, intervale care sunt imagini prin *X* ale unor evenimente din *K*.

Dacă variabila aleatoare *X* ia valori într-o multime cel mult numerabilă

$$\{x_i / x_i \in R,\ i \in I, I \subset N^+\}$$

atunci ea se numeste discretă si

$$\sum_{i \in I} P_X(x_i) = 1$$

$$P_X(J) = \sum_{x_i \in J} P_X(x_i),\ \ \forall J \in K_X$$

Dacă *X* poate lua toate valorile unui interval $I \in K_X$ atunci probabilitatea asociată intervalului este

$$P_X(I) = \int_I f_X(x)dx$$

si este o functie absolut continuă. Functia $f_X(x)$ care apare în formulă se numeste *densitatea de probabilitate* sau *densitatea de repartitie* a variabilei aleatoare *X*, este nenegativă pentru orice *x* si are proprietatea

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

Functia

$$F_X(x) = P[X(\omega) < x] = P_X[(-\infty, x)] = \int_{-\infty}^{x} f_X(x)dx$$

se numeste *functie de repartitie* a variabilei aleatoare X

Functia de repartitie este nedescrescătoare pe întreaga axă reală

$$a < b \Rightarrow F_X(a) \le F_X(b) \quad \forall a, b \in R$$

si este continuă la stânga în fiecare punct

$$\lim_{x \to a, x < a} F_X(x) = F_X(a) \ \forall a \in R$$

Valorea minimă si valoarea maximă ale unei functii de repartitie sunt date de limitele

$$\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1$$

Eventualele discontinuităti sunt de speta primă si sunt cel mult numerabile. Reciproc, orice functie cu proprietătile de mai sus poate fi pusă în corespondentă cu un câmp de probabilitate.

Pentru o variabilă aleatoare discretă

$$F_X(x) = \sum_{x_i < x} P_X(x_i)$$

iar pentru una continuă înafară de relatia scrisă deja mai sus

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx$$

are loc si relatia

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Pentru orice interval $[a, b) \subset R \quad P_X\{[a, b)\} = F_X(b) - F_X(a)$ si pentru orice valoare a, $P_X(a) = 0$.

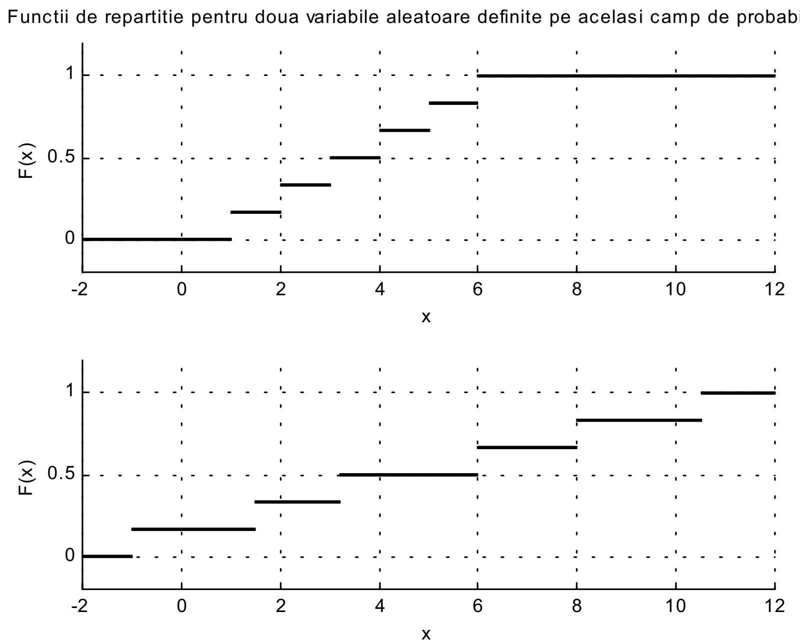
În referirea făcută putin mai devreme la cazul zarului s-a semnalat posibilitatea ca pe acelasi câmp de probabilitate să se definească nu una ci mai multe variabile aleatoare. Se notează uzual cu $V(\Omega, K, P)$ multimea tuturor variabilelor aleatoare definite pe câmpul de probabilitate trecut între paranteze.

Dacă $X, Y \in V(\Omega, K, P)$ atunci suma, diferenta, produsul celor două variabile aleatoare, modulul, puterea, în general o functie măsurabilă Borel de oricare dintre ele sunt toate variabile aleatoare din multimea $V(\Omega, K, P)$.

Ori de câte ori nu este pericol de confuzie, variabila aleatoare trecută până acum ca indice al functiei de repartitie sau al functiei densitate de probabilitate/repartitie poate lipsi din acea pozitie.

Dacă se reia exemplul zarului, care la fiecare experientă este făcut să se rostogolească pe o suprafată plană, orizontală, atunci multimea evenimentelor elementare (atomi) Ω este alcătuită din aparitiile deasupra a celor sase fete, marcate uzual cu unu până la sase puncte. Multimea de părti ale lui Ω este alcătuită din evenimentele elementare si din toate reuniunile posibile de evenimente elementare la care se adaugă evenimentul imposibil.

Multimea K organizată ca un corp de evenimente coincide chiar cu multimea de părti $P(\Omega)$, iar functia numită probabilitate ia valoarea 1/6 pentru fiecare din evenimentele elementare deoarece fetele zarului au sanse egale de a apărea deasupra.



Cum s-a mai spus, numărul de puncte observat pe fata de deasupra a zarului poate fi considerat o variabilă aleatoare. În acest caz functia de repartitie se prezintă ca în graficul superior din desenul de mai sus si este, ca pentru orice variabilă aleatoare discretă, o functie în scară.
Dar pe acelasi câmp de probabilitate se pot defini si alte variabile aleatoare. Pe câmpul asociat zarului perfect se poate imagina, de pildă, functia $X:\Omega \to R$ definită astfel

$$\begin{array}{cccccc} \omega_1 & \omega_2 & \omega_3 & \omega_4 & \omega_5 & \omega_6 \\ -1 & 1{,}5 & 3{,}2 & 6 & 8 & 10{,}5 \end{array}$$

($\omega_i \equiv A_i$, i = 1, 2, 3, 4, 5, 6) si atunci functia de repartitie se prezintă diferit, ca în graficul inferior din aceeasi figură. Asadar, multimea $V(\Omega, K, P)$ este extrem de bogată.
De variabilele aleatoare sunt legate câteva valori remarcabile. Una foarte importantă este *media* definită ca

$$M(x) = \int_{-\infty}^{\infty} x f(x) dx$$

care face parte din lista nesfârsită a momentelor de diferite ordine ale variabilei, acesta fiind momentul de ordinul unu.
Cu o relatie similară se poate calcula media unei functii $g(x)$ de variabila aleatoare x având în vedere caracterul aleator al valorilor functiei

$$M[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

si dacă în particular $g(x) = x^r$, $r \in N$ avem tocmai *momentul de ordinul r* despre care s-a amintit.
În cazul particular $g(x) = [x - M(x)]^2$ se obtine o altă valoare importantă, caracteristică variabilei aleatoare descrise de functia de repartitie $F(x)$ sau de densitatea de repartitie $f(x)$, si anume *dispersia*. Rădăcina pătrată pozitivă a dispersiei se numeste *abaterea medie pătratică* a variabilei aleatoare pentru care s-a executat calculul. Dispersia este momentul centrat de ordinul doi al variabilei aleatoare, unul din multiplele momente ale variabilei, centrate fată de medie, de ordine diferite.
Nu numai variabilele aleatoare continue au momente, medii, dispersii, ci si cele discrete. În cazul discret, formulele de calcul contin sume în locul integralelor si valorile variabilei parcurg întreaga listă de valori posibile, iar densitatea de repartitie este înlocuită de probabilitătile asociate valorilor pe care variabila le poate lua efectiv.
Câteva legi de repartitie teoretice foarte utilizate sunt prezentate pe scurt în continuare.
Legea binomială (Bernoulli) este exprimată de relatia

$$P(m) = C_n^m p^m (1-p)^{n-m}$$

cu $0 \le m \le n$ si p un număr în intervalul [0, 1]. Legea binomială este de tip discret. Variabila aleatoare este m. Are media np si dispersia $np(1-p)$. Există un model fizic legat de această lege de repartitie. Modelul îl constituie o urnă cu bile de două culori, iar evenimentele constau în rezultatele extragerii repetate a câte unei bile după care bila extrasă este reintrodusă în urnă. Variabila m reprezintă numărul bilelor de o anumită culoare din cele două, în n extrageri succesive, conform schemei cu bila returnată. Numărul p reprezintă proportia de bile de acea culoare în urnă, cu alte cuvinte probabilitatea de extragere a unei bile de culoarea respectivă.
Legea Poisson exprimată sub forma

$$P(m) = \frac{\mu^m}{m!} \exp(-\mu)$$

cu $\mu > 0$ si m natural ca variabilă aleatoare. Media variabilei m este μ, dispersia ei este, de asemenea, μ. Un modelul fizic îl reprezintă numărul dezintegrărilor radioactive, numărul de apeluri telefonice solicitate într-o centrală etc. într-un interval de timp precizat, (relativ) scurt.
Legea normală (gaussiană) care este dată de densitatea de probabiltate

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

în care m este media variabilei x si σ^2 este dispersia ei. Legea normală este considerată o lege limită pentru sumele de variabile aleatoare. Un fenomen afectat de foarte multi factori aleatori care actionează prin însumare se prezintă de cele mai multe ori ca un fenomen aleator descris de o lege normală.
Variabilele aleatoare din expunerea teoretică sau din exemplele prezentate mai sus au fost până acum simple, adică a fost vorba în toate cazurile de o singură aplicatie $X{:}\Omega \to R$ legată de un unic câmp de probabilitate (Ω, K,

P). Se pot imagina variabile aleatoare cu mai multe componente, variabile aleatoare sub forma unor vectori cu componente aleatoare definite relativ la un acelasi câmp de probabilitate sau la câmpuri de probabilitate diferite. Astfel, legea următoare se referă la o variabilă aleatoare vectorială.

Legea normală multidimensională dată de densitatea de repartitie

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det W}} e^{-\frac{1}{2}(x-m)^T W^{-1}(x-m)}$$

cu media m, un vector cu n componente, si cu matricea de covariatie W, o matrice $n\times n$ pozitiv definită. Pentru ca expresia dată să aibă consistenta necesară trebuie definită mai exact matricea W.

Este de comentat mai întâi problema corelatiei a două variabile aleatoare. Două variabile aleatoare pot fi necorelate, caz în care valorile uneia nu influentează în nici un fel valorile pe care le poate lua cealaltă, dar, alternativ, pot fi mai mult sau mai putin dependente ceea ce înseamnă că dacă una din variabile a luat o valoare atunci legea de repartitie a celeilalte se modifică în functie de acea valoare a primei variabile.

Fiind date două variabile aleatoare x si y de medii nule, media produsului lor $M(xy)$ se numeste covariatie. Dacă covariatia este nulă se poate spune în general că cele două variabile nu sunt corelate. Dimpotrivă, dacă $M(xy) \neq 0$ variabilele sunt corelate, există o corelatie între ele, există o dependentă între valorile pe care ele le iau în sensul arătat putin mai devreme. Dacă mediile sunt diferite de zero, afirmatia si definitia se mentin pentru abaterile de la medie. Întrucât covariatia $M(xy)$ poate lua valori foarte diferite, pentru o apreciere cantitativă mai riguroasă a tăriei corelatiei se utilizează coeficientul de corelatie

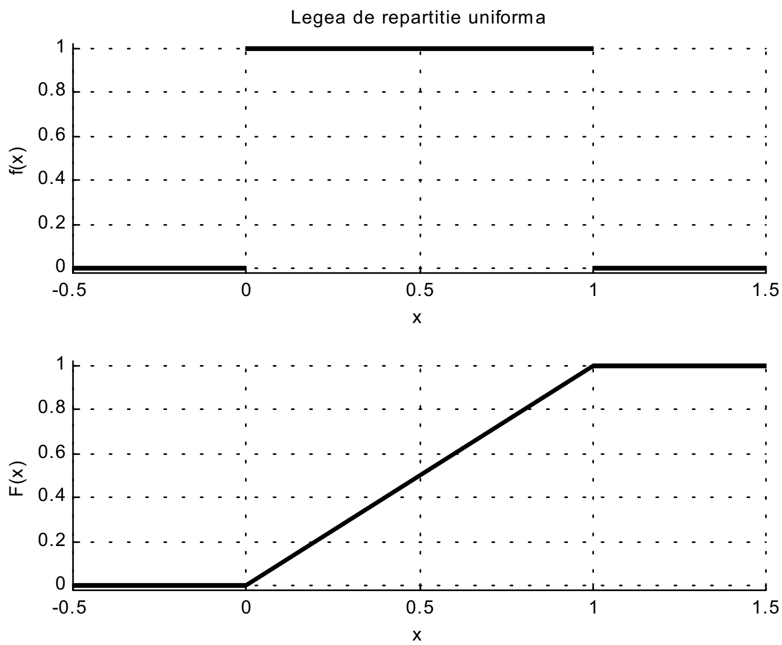
$$\rho = \frac{M(xy)}{\sqrt{M(x^2)\,M(y^2)}}$$

care ia valori în intervalul [–1, 1] si în expresia căruia se disting dispersiile celor două variabile, $M(x^2)$ si $M(y^2)$. O valoare apropiată de extremele intervalului indică o corelatie strânsă, o valoare apropiată de zero exprimă o corelatie slabă.

Componentele unui vector aleator, privite ca variabile aleatoare simple sunt mutual mai mult sau mai putin corelate. Se defineste ca matrice a covariatiilor unui vector aleator x media produsului xx^T, adică media produsului vectorului cu transpusul său. Se obtine o matrice pătrată simetrică care are pe diagonală dispersiile componentelor pure. Aceasta este matricea W utilizată în expresia densitătii de repartitie a variabilei aleatoare normale multidimensionale din discutia de mai sus. Dacă matricea covariatiilor este diagonală (are toate elementele nule cu exceptia celor de pe diagonala principală) atunci componentele vectorului aleator sunt mutual independente. Împărtirea fiecărui element al matricei covariatiilor cu abaterile medii pătratice ale componentelor corespunzătoare ale vectorului x produce o matrice a coeficientilor de corelatie, cu 1 pe diagonală, cu valori in intervalul [–1, 1] în rest.

Generarea de numere aleatoare

În modelarea si mai ales în simularea sistemelor este necesară deseori generarea de numere aleatoare a căror aparitie să se producă conform unei anumite legi de repartitie: unele valori mai frecvent, altele, eventual, mai putin frecvent. În sprijinul acestei cerinte, aproape toate limbajele de programare evoluate au în biblioteca lor, alături de alte functii, functii generatoare de numere aleatoare uniform repartizate pe un interval precizat, de regulă intervalul (0, 1). În PASCAL, de pildă, există functia **`random`**, cu sau fără argument, care generează astfel de numere. Subprogramul **`randomize`** invocat înaintea primului apel la functia **`random`** asigură secvente de numere diferite la fiecare nouă utilizare succesivă într-un program a functiei de bibliotecă generatoare de numere aleatoare. Versiunea fără argument a functiei **`random`** produce numere reale în intervalul (0, 1), uniform repartizate pe acel interval.



Versiunea cu argument de tip **`word`**, **`random(w)`**, produce numere aleatoare de tipul **`word`**, cuprinse între 0 si $w - 1$. Pentru cazul continuu al functiei **`random`** fără argument, functia densitate de repartitie si functia de repartitie sunt figurate în graficele alăturate.
Cu generatorul de numere aleatoare **`random`** sau cu generatoarele similare din alte limbaje se pot genera numere aleatoare repartizate după alte legi, diferite de cea uniformă. Pentru aceasta se pot utiliza metode analitice sau o metodă directă care are în vedere functia de repartitie a variabilei care trebuie generată.
Legea de repartitie normală *normată* (de medie nulă si de dispersie 1) este legată de legea de repartitie uniformă prin una sau alta dintre relatiile următoare

$$u_1 = \sqrt{-2\ln x_1}\ \cos(2\pi\ x_2)$$
$$u_2 = \sqrt{-2\ln x_1}\ \sin(2\pi\ x_2)$$

în care x_1 si x_2 sunt două numere aleatoare independente, cu repartitie uniformă pe intervalul (0, 1). Este un exemplu de generare analitică a unor numere aleatoare supuse unei legi de repartitie diferită de cea uniformă.
Un alt exemplu este cel al generării de numere aleatoare uniform repartizate pe un interval finit (a, b) oarecare. Trecerea la noua variabilă se realizează prin mijlocirea relatiei

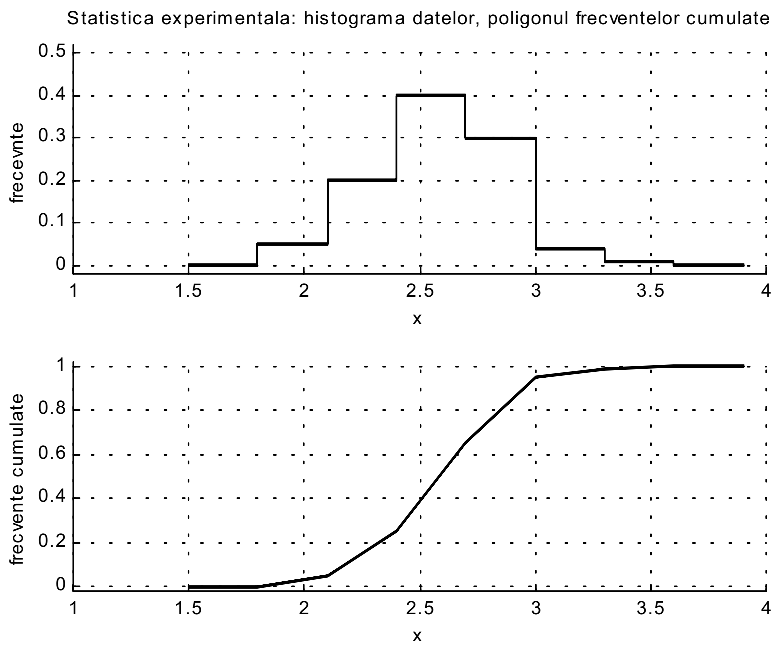
$$u = a + (b - a)x$$

cu x generat de functia de bibliotecă **`random`**. Variabila u este uniform repartizată pe intervalul finit specificat.
Varianta analitică de generare a unor numere aleatoare repartizate conform unei legi particulare nu este totdeauna posibilă. Modul de generare alternativ este descris în continuare.
Se admite că este dată functia de repartitie $F(u)$ a unei variabile u sau functia ei densitate de repartitie $f(u)$ din care se poate calcula $F(u)$. Se generează valori x uniform repartizate pe intervalul (0, 1) cu ajutorul functiei de bibliotecă **`random`** sau similara ei din alte limbaje de programare. Se calculează de fiecare dată $u = F^{-1}(x)$, unde $F^{-1}(.)$ este inversa functiei de repartitie a variabilei u care trebuie generată. Functia de repartitie este totdeauna o functie monotonă, deci este inversabilă pentru orice $x \in (0,1)$. Intervalul (0, 1) este multimea de valori comună tuturor functiilor de repartitie. Variabila aleatoare u este cu sigurantă repartizată conform legii date de functia $F(u)$ sau de derivata ei $f(u)$.

Raportul experiment-lege de repartitie teoretică

Variabilele aleatoare pot fi observate prin valorile pe care ele le iau efectiv. Numărul observatiilor este inevitabil finit. Forma matematică a legii de repartitie precum si parametrii ei, inclusiv cei mai simpli cum sunt media si dispersia, sunt elemente care trebuie *inferate*, obtinute prin *inferentă* din aceste observatii experimentale. *Inferenta* este operatia logică prin care se admite o judecată al cărui adevăr nu este verificat direct ci în virtutea unei legături a ei cu alte judecăti considerate adevărate.
Cu toate că ipoteza normalitătii unei variabile aleatoare este satisfăcătoare în foarte multe cazuri, în special atunci când fenomenul este rezultatul actiunii întâmplătoare a unui număr mare de factori, uneori reprezentativitatea legii de repartitie însăsi trebuie verificată. Verificarea se face, desigur, pe baza unui volum limitat de observatii experimentale.
Observatiile experimentale, fie acestea $x_1, x_2, ..., x_n$, sunt mai întâi sortate pe m intervale I_k $(k = 1,2,...,m)$ în care este partitionată convenabil axa reală. Sortarea se face în raport cu apartenenta lor la unul sau altul din acele intervale. Se calculează frecventele absolute n_k $(k = 1,2,...,m)$ pentru fiecare interval adică numărul de valori observate care apartin unuia sau altuia din intervalele I_k. Cu aceste frecvente sau cu frecventele relative

obtinute prin împărtirea lor la numărul total de valori observate *n* se poate trasa un grafic sub forma unor dreptunghiuri cu baza cât fiecare interval si înăltimea egală cu frecventa. Aceste grafice sunt denumite *histograme* ale frecventelor relative sau absolute. Prin cumularea ordonată a frecventelor se obtine un grafic numit *poligonul frecventelor* relative sau absolute cumulate.



Cele două functii grafice sunt pentru colectia de date experimentale ceea ce pentru variabila aleatoare sunt probabilitătile sau densitatea de repartitie si functia de repartitie. În termeni de frecvente relative, functiile în formă grafică care au ca sursă experimentul ar trebui să estimeze functiile teoretice corespondente. Dacă ele sunt sau nu estimatii ale acelor functii teoretice, dacă legea de repartitie teoretică reprezintă într-adevăr variabila aleatoare observată se apreciază prin calculul unei valori

$$\chi^2 = \sum_{k=1}^{m} \frac{(n_k - np_k)^2}{np_k}$$

în care intervin atât frecventele experimentale cât si probabilitătile teoretice $p_k = P(x \in I_k)$, $(k = 1,2,\ldots,m)$ si care este o variabilă aleatoare deoarece, evident, la un nou set de observatii se obtine aproape sigur altă valoare.
Variabila χ^2 este consacrată în statistica matematică si este definită ca o sumă de pătrate ale unor variabile aleatoare normale normate (de medie nulă si de dispersie egală cu unitatea). Variabila are un număr de *grade de libertate* egal cu numărul de termeni din suma definitorie. Functia de repartitie a variabilei χ^2 este tabelată sau se poate evalua numeric si este folosită în verificarea ipotezelor statistice de natura celei formulate mai devreme sau de altă natură.
Intuitiv, valoarea χ^2 calculată din observatii experimentale ar trebui să fie cât mai apropiată de zero. Atunci, probabilitătile p_k ar fi foarte apropiate de

frecventele relative n_k / n rezultate din experiment. Ipoteza *modelul teoretic este verificat de observatiile experimentale* (H_0) se opune ipotezei alternative (H_1) *modelul teoretic pe cale de a fi adoptat nu este verificat de observatiile experimentale*. Pragul discriminator între cele două adevăruri mutual exclusive este stabilit ca limită superioară a unui interval definit pentru un nivel de semnificatie sau pentru un nivel de încredere precizat q (uzual 0,95), interval care grupează 100q% din valorile χ^2 firesti, plauzibile în cazul valabilitătii ipotezei H_0. Schema acceptării (sau respingerii) uneia sau alteia dintre ipoteze este

$$\chi^2 \overset{H_0}{\underset{H_1}{\lessgtr}} \chi_q^2$$

cu valoarea de provenientă experimentală în stâga semnului discriminator, cu valoarea teoretică (tebelară) în dreapta acelui semn. Pe calea aceasta se poate selecta legea de repartitie adecvată.

O aplicatie la fiabilitatea sistemelor de productie

Problema sigurantei în functionare a unei masini, a unui agregat, a unui sistem în general, functionare raportată la îndeplinirea unui obiectiv tehnologic precizat este de foarte mare importantă economică atât pentru producătorul acelor sisteme cât si pentru utilizatorul lor.
Producătorul trebuie să-si organizeze un serviciu în ajutorul clientului cel putin pentru perioada de garantie si să dimensioneze sitemul de service corespunzător cu parametrii de fiabilitate ai sistemelor pe care le produce. Utilizatorul, în cazul în care isi asumă mentinerea în functiune a respectivelor sisteme după perioada de garantie prin forte proprii trebuie să cunoască elemente precum durata (aleatoare) pâna la prima (care poate fi si ultima în unele cazuri) defectare sau durata (tot aleatoare, desigur) între două defectări succesive pentru a-si dimensiona cât mai bine echipa de întretinere, pentru a face o aprovizionare cu piese înlocuitoare ratională.
Un studiu de fiabiliate la producător este necesar. Distributia duratei de viată a unui echipament, care este o variabilă aleatoare se poate uneori deduce din carateristicile similare ale pieselor sau subansamblelor componente ale acelui echipament. În conditii de noutate totală a produselor, producătorul trebuie să facă propriile sale studii de fiabilitate.
Observarea sub aspectul fiabilitătii a mai multor produse identice conduce uzual la acumularea unor durate de functionare experimentale. Aceste durate, fie ele în număr de n, se pot folosi la aprecierea reprezentativitătii unei anumite legi de repartitie.
Dacă axa timpului este împărtită în m intervale, aceste durate pot fi sortate/numărate pentru fiecare din aceste intervale obtinându-se frecventele n_k si frecventele relative n_k / n pentru fiecare interval $I_k (k = 1,2,...,m)$. Dacă densitatea de repartitie avută în vedere este $f(t)$ atunci se pot calcula probabilitătile

$$p_k = \int_{I_k} f(t)dt$$

asociate fiecărui interval, valori care sunt, desigur, teoretice. Frecventele relative sunt estimări experimentale ale acestor probabilităti. Se constată, desigur, diferente între probabilităti si estimările lor. Aceste diferente pot servi la formularea unor ipoteze privind adecvarea modelului teoretic la experimentul observat.

O modalitate de decizie asupra acestei adecvări se bazează pe retinerea diferentei celei mai importante ca valoare absolută si compararea ei cu anumite tabele care dau norme în ceea ce priveste abaterea maximă îngăduită. Este vorba aici de utilizarea testului Kolmogorov-Smirnov.

O altă posibilitate mai larg utilizată este aceea care folseste o variabilă χ^2, asa cum s-a arătat mai sus.

Estimarea si verificarea parametrilor legii de repartitie teoretice

Asa cum s-a arătat, legea de repartitie cea mai frecvent utilizată în modelarea si simularea dinamicii sistemelor este legea normală. Ipotezele si verificările parametrice discutate în cele ce urmează se referă în exclusivitate la variabile aleatoare repartizate normal sau, cum se mai spune, gaussian.

Selectie, parametri de selectie

O listă de valori observate $x_1, x_2, ..., x_n$ ale unei variabile aleatoare poartă numele consacrat de *selectie*. Numele ar putea părea impropriu prin prisma sensului uzual al cuvântului *selectie* si, de aceea, trebuie subliniat că valorile din lista de observatii nu comportă nici un proces subiectiv de alegere. Prin selectie se întelege numai retinerea experimentală a unui număr finit de valori ale variabilei aleatoare din numărul extrem de mare de valori pe care aceasta le poate lua.

Se admite că variabila se distribuie normal cu media μ si dispersia σ^2.

Se defineste ca *medie de selectie* media aritmetică a valorilor observate

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Ea este o estimatie absolut corectă a mediei μ si se repartizează ca si variabila x observată, normal, cu aceeasi medie μ dar cu dispersia σ^2 / n.

Din datele care alcătuiesc selectia se pot calcula două *dispersii de selectie*

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

si

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Ambele sunt estimări absolut corecte ale dispersiei σ^2, prima cu *n* grade de libertate, a doua mai uzuală deoarece nu necesită cunoasterea mediei teoretice μ, cu $n - 1$ grade de libertate.
Sintagma *estimatie absolut corectă* exprimă faptul că media unei astfel de etimatii este exact valoarea parametrului estimat.

Ipoteze asupra parametrilor unei legi de repartitie normale

Asupra mediei, ipoteza cea mai frecventă are forma $\mu = \mu_0$ si este notată cu H_0. Opusa ei se notează cu H_1. Uneori se formulează ipoteze în foma unilaterală, în care semnul egal este înlocuit cu un semn de inegalitate. Verificarea uneia sau alteia dintre ipoteze se face în baza unei selectii.
Dacă dispersia σ^2 este cunoscută atunci se poate calcula o valoare

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

care este o variabilă aleatoare repartizată normal si este în plus si normată adică are media nulă si dispersia egală cu unitatea. Se stabileste un nivel de încredere sau de semnificatie *q*, uzual de 0,95 (95%), pentru care în tabele sau prin calcul se găseste un z_q, în particular $z_{0.95} = 1.96$. Această valoare delimitează un interval $(-z_q, +z_q)$, numit interval de încredere, care contine în mod normal 95% din valorile *z* calculate din datele selectiilor de tipul si de volumul specificat mai devreme. Prin urmare valorile din afara intervalului de încredere sunt cu totul improbabile, probabilitatea lor fiind complementara $1 - q$. Aparitia unei valori *z* din această din urmă categorie pune sub semnul îndoielii valabilitatea ipotezei formulate. Asadar, dacă valoarea *z* calculată apartine intervalului de încredere, $z \in (-z_q, +z_q)$, ipoteza se acceptă, în caz contrar se respinge sau, într-o exprimare care pune în evidentă ambele ipoteze, mutual exclusive

$$|z| \underset{H_1}{\overset{H_0}{\lessgtr}} z_q$$

Dacă dispersia σ^2 nu este cunoscută se utilizează variabila Student, în care intervine radicalul pozitiv al dispersiei de selectie

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Variabila Student este o variabilă aleatoare în legătură cu care se mentionează si un număr de grade de libertate, acelasi cu al estimatiei s^2.

Se stabileste un nivel de încredere q si un interval de încredere $(-t_q, +t_q)$. Tabelele dau aceste valori pentru diverse niveluri de încredere, cel mai uzual fiind acelasi 0,95, dar si pentru grade de libertate diferite. Ipoteza H_0 se confirmă dacă valoarea Student calculată se situează în interiorul intervalului de încredere. Ipoteza se respinge în caz contrar. Sintetic

$$|t| \underset{H_1}{\overset{H_0}{\lessgtr}} t_q$$

Ipoteze se fac, de asemenea si asupra dispersiei. Ipotezele acestea trebuie si ele verificate. Astfel, fiind date două estimatii ale aceleiasi dispersii s_1^2 si s_2^2, valoarea bazată pe experiment

$$F = \frac{s_1^2}{s_2^2}$$

are cracteristicile unei variabile aleatoare si se spune că are f_1 si f_2 grade de libertate, respectiv gradele de libertate ale celor două estimatii raportate. Variabila este cunoscută in statistica matematică sub numele de variabila F (Fisher-Snedecor). În particular, gradele de libertate pot fi f_1 si ∞ si atunci

$$F = \frac{s_1^2}{\sigma^2}$$

este o variabilă F cu f_1 si ∞ grade de libertate. O ipoteză $\sigma^2 = \sigma_0^2$ se poate verifica prin calcularea unei valori F cu σ_0^2 la numitor. Un nivel de semnificatie q delimitează si în acest caz un interval de încredere. Un F calculat din date experimentale superior lui F_q tabelar cu gradele de libertate respective impune respingerea ipotezei H_0 formulate si acceptarea ipotezei alternative H_1. Cazul contrar face ca ipoteza H_0 să fie acceptată. Schema globală este cuprinsă în exprimarea

$$F \underset{H_1}{\overset{H_0}{\lessgtr}} F_q$$

Criteriul F se aplică de obicei unilateral.

Variabila z normală normată si variabilele aleatoare t si F care sunt în conexiune directă cu legea de repartitie normală permit formularea si testarea unui număr important de ipoteze statistice. O altă variabilă aleatoare importantă legată de variabila repartizată normal este variabila χ^2 despre care s-a vorbit la verificarea calitătii de model al unei variabile aleatoare observate experimental, îndeplinită de o lege de repartitie teoretică. Variabila χ^2 este o sumă de pătrate ale unor variabile normale normate independente si are gradele de libertate egale cu numărul de termeni în sumă. Variabila χ^2 permite ea însăsi verificarea de ipoteze asupra dispersiei dat fiind faptul că într-o estimatie s^2 a dispersiei teoretice σ^2 se poate separa o sumă de pătrate de variabile z normale normate independente si, implicit, o variabilă χ^2. Tabele sau calculul direct dau si în cazul acesta valori χ_q^2 care constituie pragul discriminator de ipoteze la un nivel de semnificatie q precizat.

Prelucrarea datelor experimentale

Modele matematico-statistice

Prin observarea concomitentă a două sau mai multe variabile afectate de componente aleatoare se poate stabili nu numai o corelatie calitativă dată de un coeficient de corelatie cum s-a explicat mai devreme ci se pot evalua uneori corelatii foarte asemănătoare dependentelor functionale. Operatia este cunoscută sub denumirea generică de *modelare statistico-matematică*. Uzual forma legăturii functionale este cunoscută. Se pune numai problema ca pornind de la o listă de observatii experimentale să se estimeze parametrii din acea functie, care fac "acordul" functiei cu rol de model pe datele experimentale la dispozitie. Evaluarea acestor parametri, întocmai ca evaluarea unor parametri ai legilor de repartitie normale discutată mai sus face parte dintr-un cadru mai larg de probleme cunoscut sub numele generic de *estimarea parametrilor*. Este o estimare pentru că parametrii au încă un caracter aleator. Relatiile-model stabilite pe calea estimării de parametri pot fi utilizate în calcule tehnice diverse.

Estimarea de parametri în relatii-model algebrice

Un model algebric are în general forma

$$y = f(x, a)$$

cu f o functie vectorială (mx1) de vectorul de variabile x (nx1) care contine vectorul de parametri a (px1).

Problema estimării parametrilor se pune în termenii următori: fiind dată o listă de perechi de valori experimentale x si y se cere a se determina parametrii a astfel încât să fie minimizat un criteriu de distantă dintre model si experiment. Dintre criteriile posibile sunt frecvent utilizate cele mai mici pătrate cu sau fără ponderi, suma abaterilor luate în valoare absolută ca atare sau raportate la modulul valorilor experimentale. În toate aceste alegeri distanta dintre model si experiment se referă numai la valorile y cu acceptarea ipotetică, tacită sau explicită, a unei precizii mult mai bune în măsurarea variabilelor x decât în observarea lui y. Uneori însă variabilele independente x sunt afectate ele însesi de erori de observare si de mentinere la anumite valori în cursul experimentelor, care nu pot fi ignorate. În cazul acesta în evaluarea acelei distante model-experiment intră si variabilele x după cum se va explica mai departe.

Metoda celor mai mici pătrate

Parametrii a din relatia

$$y = f(x,a)$$

pot fi determinati din date experimentale având în vedere că în realitate relatia este îndeplinită sub forma aproximativă

$$y = f(x,a) + \varepsilon$$

Scrisă pentru mai multe puncte experimentale

$$y_k = f(x_k,a) + \varepsilon_k \ (k = 1,2,...,N)$$

aceasta permite constituirea unui criteriu de distantă între model si experiment de forma

$$S = \varepsilon^T Q \varepsilon$$

cu ε vectorul rezidualelor ε_k si cu Q o matrice de ponderi pozitiv definită. Matricea Q este de cele mai multe ori diagonală si dă ponderi diferite unor observatii y_k afectate de erori variabile cu k. Dacă erorile sunt constante si sunt descrise statistic de o aceeasi lege de repartitie admisă a fi normală de medie nulă, atunci matricea Q poate fi matricea unitate I multiplicată eventual cu valoarea reciprocă a dispersiei unice, caz în care metoda coincide cu *metoda celor mai mici pătrate* clasică, cu ponderi constante pentru cele N observatii experimentale, de fapt fără ponderi. Parametrii a căutati sunt aceia care minimizează pe S, care este o sumă de pătrate ale abaterilor model-experiment, ponderate sau nu. Cazul cel mai frecvent în aplicatii si în consecintă cel mai pus la punct sub aspect teoretic este cel liniar în parametrii a, cu alte cuvinte cel în care coeficientii apar o singură dată fiecare la puterea întâia. Aparent particular, cazul devine destul de general dacă se iau în consideratie posibilitătile de liniarizare fie prin substitutii adecvate fie prin dezvoltări Taylor valabile pe regiuni limitate ale spatiului x. Prin urmare, merită o atentie aparte cazul liniar

$$y = x^T a$$

în care vectorul x contine uzual si o primă componentă constantă si egală cu unitatea, care corespunde coeficientului liber de orice influentă datorată modificărilor lui x. Vectorul a al parametrilor este $(n + 1)$-dimensional adică are n componente, câte una pentru fiecare componentă variabilă a vectorului x si încă una ca termen liber.

Dacă y_k sunt valorile observate si x_k sunt valori particulare ale vectorului x în experientele sau observatiile $k = 1,2,...,N$, atunci minimul sumei S se obtine pentru solutia sistemului în coeficientii necunoscuti a

$$X a = Y$$

solutie în sensul celor mai mici pătrate. În relatia ultimă, matricea X are ca linii vectorii x_k^T, iar Y este vectorul observatiilor y_k. Sistemul este liniar în componentele lui a si se poate rezolva, în etape, prin premultiplicarea mai întâi cu transpusa matricei X

$$X^T X a = X^T Y$$

si, după aceea, prin multiplicarea la stânga cu inversa matricei produs $X^T X$

$$a = (X^T X)^{-1} X^T Y$$

Matricea $(X^T X)^{-1} X^T$ mai este numită si inversa generalizată sau pseudoinversa matricei X, dacă inversa matricei $X^T X$ există.

Existenta inversei utilizate este un mod de a vorbi despre diversitatea punctelor x_k. Desigur, în matricea X se pot încorpora valori ale vectorului x variate, asa cum rezultă din observarea curentă a sistemului de modelat. Este vorba în acest caz de varianta experimentului pasiv, nedirijat. Experimentul se poate însă planifica, în primul rând pentru a asigura acea diversitate de valori ale componentelor vectorului x capabilă să pună în evidentă efectele lor asupra valorilor y. Planificarea poate merge încă mai departe prin alegerea componentelor vectorului x în asa fel încât să aibă loc relatia

$$\sum_{k=1}^{N} x_{ik} x_{jk} = 0$$

pentru oricare două componente distincte $i \neq j$. De pildă, experimentul din tabelul următor

Experienta nr.	x_0	x_1	x_2
1	1	–1	–1
2	1	–1	1
3	1	1	–1
4	1	1	1
5	1	0	0

are această proprietate, numită proprietatea de *ortogonalitate*. Pare dificil de pus în aplicare un asemenea plan experimental. Dar în tabelul de mai sus nu este vorba de valori *naturale* ale variabilelor ci de valori legate într-un mod adecvat de cele naturale. Mai explicit, dacă variabilele din realitate sunt, de pildă, o temperatură T si un debit d, care în cursul experimentării iau valorile 50, 75, 100 °C, respectiv 1000, 1200, 1400 kg/oră atunci variabilele

$$x_1 = \frac{T - 75}{25}$$

$$x_2 = \frac{d - 1200}{200}$$

iau exact valorile din tabel. Coeficientii relatiei-model se estimează în raport cu aceste variabile si apoi se revine la variabilele T si d din realitate.

Avatajul unui experiment planificat si, în plus, ortogonal este dublu. Pe de o parte matricea $X^T X$ este diagonală si deci usor de inversat. Pe de altă parte efectul fiecărei variabile, usor de calculat

$$a_l = \frac{\sum_{k=1}^{N} x_{lk} y_k}{\sum_{k=1}^{N} x_{lk}^2} \quad (l = 0,1,2,\ldots,n)$$

poate fi pus în evidentă separat, în deplina lui semnificatie (sau lipsă de semnificatie). În aceste conditii suma de pătrate ale rezidualelor (un alt termen pentru diferentele dintre valorile experimentale y_k si cele calculate cu relatia $y = x^T a$ în acelasi conditii x_k) se descompune sub forma

$$\sum_{k=1}^{N}(y_k - a_0x_{0k} - a_1x_{1k} - ... - a_nx_{nk})^2 =$$

$$= \sum_{k=1}^{N} y_k^2 - a_0^2\sum_{k=1}^{N} x_{0k}^2 - a_1^2\sum_{k=1}^{N} x_{1k}^2 - ... - a_n^2\sum_{k=1}^{N} x_{nk}^2$$

care rearanjată duce la

$$\sum_{k=1}^{N} y_k^2 = a_0^2\sum_{k=1}^{N} x_{0k}^2 + a_1^2\sum_{k=1}^{N} x_{1k}^2 + ... + a_n^2\sum_{k=1}^{N} x_{nk}^2 +$$

$$+ \sum_{k=1}^{N}(y_k - a_0x_{0k} - a_1x_{1k} - ... - a_nx_{nk})^2$$

Această ultimă expresie pune în evidentă o interesantă descompunere a sumei pătratelor valorilor observate y_k, din partea stângă a egalitătii. Descompunerea contine termeni legati clar de câte un efect al uneia sau alteia dintre variabile si un ultim termen care este însăsi suma rezidualelor ridicate la pătrat, minimizată. Cu terminologia *sume de pătrate* si *grade de libertate* asociate, termenii din dreapta semnului de egalitate au fiecare câte 1 grad de libertate, cu exceptia ultimului care are $N - (n + 1)$ grade de libertate, adică diferenta dintre numărul de observatii asupra lui y si numărul de coeficienti cuprinsi în vectorul a. Dacă nu există nici un efect real, adică semnificativ, al x-ilor asupra lui y atunci se poate considera că valorile $a_0, a_1, ..., a_n$ sunt datorate exclusiv zgomotului (termen care denotă erorile care însotesc obisnuit măsurătorile) care acompaniază observatiile y si atunci termenii sumei de mai sus pot servi la calculul unor estimatii cu 1 sau $N - (n + 1)$ grade de libertate ale dispersiei valorilor y. Cu aceste estimatii se pot calcula valori F (Fisher-Snedecor) cu gradele de libertate respective. De pildă raportul

$$F = \frac{a_l^2\sum_{k=1}^{N} x_{lk}^2}{\dfrac{\sum_{k=1}^{N}(y_k - a_0x_{0k} - a_1x_{1k} - ... - a_nx_{nk})^2}{N-(n+1)}}$$

este o valoare F cu 1 si $N - (n + 1)$ grade de libertate. Selectând un nivel de semnificatie q (uzual 0,95) tabelele indică o valoare F_q critică. Un F calculat care se situează sub valoarea critică arată că efectul variabilei l nu există sau, în alti termeni, nu este semnificativ, valoarea a_l calculată fiind datorată exclusiv zgomotului. Dimpotrivă, un F calculat care este mai mare decât valoarea critică indică un efect semnificativ: variabila y depinde efectiv de x_l. Termenul legat de reziduale poate fi el însusi testat ca semnificatie dacă este raportat la o estimare a dispersiei din puncte experimentale repetate în aceleasi conditii. Pentru fiecare punct x repetat astfel se calculează o dispersie s^2. Se calculează apoi un s^2 global ca o medie ponderată cu gradele de libertate ale estimatiilor punctuale. Acest din urmă s^2 are ca grade de libertate suma gradelor de libertate ale estimatiilor s^2 punctuale componente. Se poate calcula acum un F ca raport al estimatiei din reziduale la estimatia din experiente repetate. Dacă acesta este sub valoarea critică tabelară se constată o situatie normală: modelul reprezintă experimentul în limitele zgomotelor care afectează măsurătorile *y*. Din contra, depăsirea valorii critice dezvăluie relatii $y \leftrightarrow x$ mai complicate, efecte ignorate cu voie sau fără voie, sau o inadecvare de altă natură a modelului la experimentul sau obiectul modelat. Valoarea *F* calculată în acest mod, sau numai termenul din suma pătratelor asociat rezidualelor când calculul unui *F* nu este posibil se poate constitui în criteriu de discriminare între două sau mai multe relatii-model posibile pentru un acelasi obiect, pe baza acelorasi date experimentale. Este de preferat aproape totdeauna modelul cu *F* mai mic, asadar cu reziduale mai mici. Desigur, variabilele care pot fi modificate după dorintă într-un experiment planificat sunt cele de intrare, independente.

Există o variabilă deosebită, timpul, care este mai putin planificabilă. Cel mai curent mod de a trata timpul în observatiile experimentale este de al măsura si marca la intervale regulate. Pe o secventă de momente egal distantate este posibilă o ortogonalizare a matricei *X* prin utilizarea unor polinoame ortogonale pe multimea de puncte de pe axa timpului. Asta presupune că sunt de calculat dependente de timp polinomiale, reprezentabile prin polinoame. Există polinoame de grad 0, 1, 2 etc. care pe o retea de puncte echidistante $\{t_0, t_1, ..., t_N\}$ au proprietatea importantă

$$\sum_{k=0}^{N} P_i(t_k) P_j(t_k) = 0$$

ori de câte ori gradele lor *i* si *j* sunt diferite. Aceste polinoame au expresiile următoare

$$P_0(t) = 1.0$$

$$P_1(t) = 1.0 - 2\frac{t}{N}$$

$$P_2(t) = 1.0 - 6\frac{t}{N} + 6\frac{t(t-1)}{N(N-1)}$$

$$P_3(t) = 1.0 - 12\frac{t}{N} + 30\frac{t(t-1)}{N(N-1)} - 20\frac{t(t-1)(t-2)}{N(N-1)(N-2)}$$

....

$$P_m(t) = \sum_{k=0}^{m} (-1)^k \frac{(m+k)!}{(m-k)!k!k!} \frac{t^{(k)}}{N^{(k)}}$$

în care s-a notat $x^{(k)} = x(x-1)(x-2)...(x-k+1)$.

Multe functii pot fi aproximate prin polinoame de acest tip si în general prin polinoame. Pornind de la gradul zero, adăugând treptat câte un polinom cu grad mai mare cu o unitate fată de etapa precedentă se pot calcula coeficienti ai unui polinom-model global. Semnificatia acestor coeficienti poate fi judecată separat.

Forma unor semnale (variatii în timp) primite de la (observate la) un sistem de productie poate fi modelată prin asemenea relatii polinomiale, deduse din esantioane prelevate la intervale de timp echidistante.

Utilizări ale modelelor de natură statistică

Modelele parametrice, adică relatiile între diferite variabile tehnologice stabilite prin metode statistice sunt utile în evaluări ale comportării sistemului în conditii diferite de acelea care au servit la stabilirea relatiilor-model. Aceste evaluări pot fi de interpolare ori de câte ori noua combinatie de conditii este situată în zona unde sunt localizate si punctele care au servit la calculul relatiilor-model. Dar pot fi utilizate si la extrapolări dacă aceleasi conditii noi sunt situate în afara domeniului efectiv explorat. Extrapolările trebuie făcute cu prudentă. De la interpolări nu se asteaptă niciodată valori sigure, certe; rezultă uzual valori foarte probabile care nu exclud realizarea practică a altori valori ale variabilelor dependente apropiate însă de cele calculate. Interpolarea are rolul de a filtra semnificativul de nesemnificativ, de a face o utilă netezire a datelor.
Dacă variabila principală este timpul atunci o modelare permite elaborarea unor *prognoze*, ceea ce corespunde în timp extrapolărilor relative la variabilele de altă natură decât temporală. Prognozele au dedicat un capitol special în această lucrare.

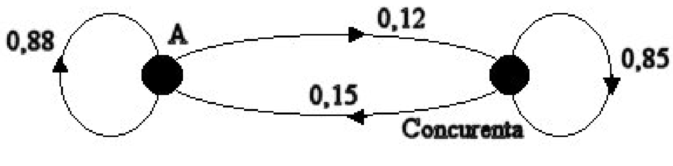
PROCESE MARCOV

O problemă tipică

Iată enuntul unei probleme: compania A, producătoare a unui tip de cereale pentru micul dejun are circa 25% din piată. Datele din anul precedent arată că 88% dintre clientii companiei au rămas atasati produsului ei, dar 12% au schimbat pentru produsul similar al concurentei. Se stie, de asemenea, că 85% din clientii concurentei au rămas loiali dar 15% au trecut la consumul produsului companiei A. Admitând stabilitatea acestor tendinte, să se determine partea de piată a companiei A în 2 ani si pe termen lung.
Această problemă este un exemplu de problemă de schimbare de brand, problemă care apare adesea în legătură cu bunurile de consum. Pentru a rezolva o problemă de acest gen se recurge la *lanturile* (*procesele*) *Markov*, care sunt un gen aparte de *procese stochastice*. Procedura în cazul particular enuntat este dată mai departe.

Procedura de obtinere a solutiei

An de an, un client poate cumpăra fie produsul A, fie produsul concurent. Pe baza acestei observatii se poate construi diagrama de mai jos unde cele două cercuri reprezintă două *stări* ale clientului generic, statistic, iar arcele reprezintă *tranzitiile* (anuale), cu probabilitătile ca un client să comute de la un produs la altul sau rămână fidel unuia dintre produse. Diagrama aceasta este cunoscută ca *diagrama de tranzitie a stărilor*. De notat că arcele sunt toate orientate.
Dată fiind această diagramă se poate construi o *matrice de tranzitie*, notată uzual cu P, care contine probabilitătile de a avea loc o tranzitie de la o stare la alta.



Punând starea 1 = consumul de produs A si starea 2 = consumul de produs al concurentei, matricea de tranzitie pentru problema formulată este

$$P = \begin{bmatrix} 0{,}88 & 0{,}12 \\ 0{,}15 & 0{,}85 \end{bmatrix}$$

cu numerele înscrise exprimând probabilitătile de tranzitie de la starea "indicele-liniei" la starea "indicele-coloanei". Cum se observă, suma elementelor pe oricare din liniile matricei de tranzitie este egală cu unitatea.

Se cunoaste că firma A detine pentru produsul ei 25% din piată. Acest fapt se concretizează într-un vector linie care reprezintă starea initială

$$s_1 = [0{,}25 \quad 0{,}75]$$

adică 25% din piată pentru produsul A si 75% pentru produsul concurent.

Teoria lui Markov spune că în perioada (anul) t starea sistemului este dată de vectorul s_t dat de

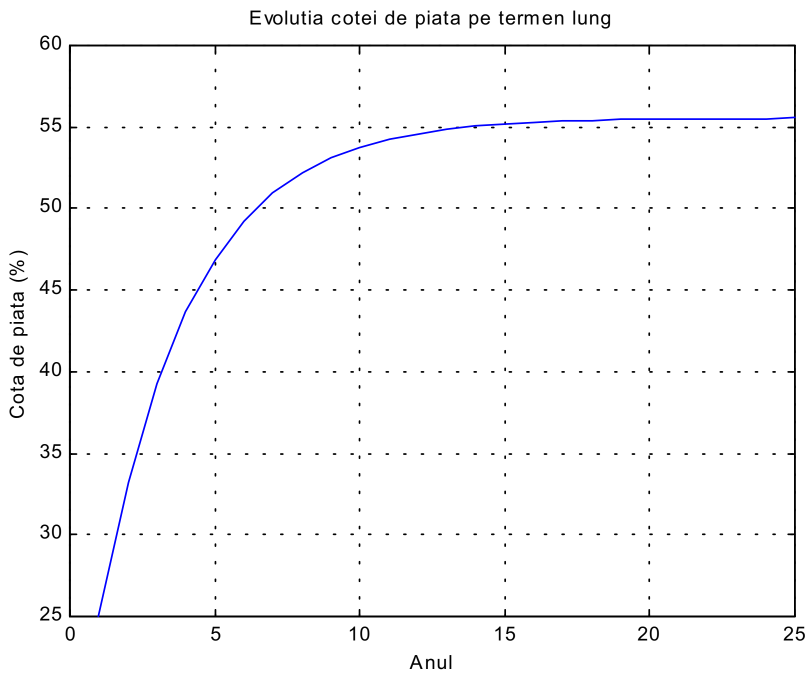
$$s_t = s_{t-1}(P) = s_{t-2}(P)(P) = ... = s_1(P)^{t-1}$$

Pentru a evalua s_t se poate evalua direct puterea a $t - 1$ a matricei P. O alternativă este a calcula starea sistemului în ani succesivi 1, 2, …, t.

Se cunoaste deja starea sistemului pentru anul 1 (s_1). Starea sistemului în anul 2 este dată de

$$s_2 = s_1 P = \begin{bmatrix} 0{,}25 & 0{,}75 \end{bmatrix} \begin{bmatrix} 0{,}88 & 0{,}12 \\ 0{,}15 & 0{,}85 \end{bmatrix} = \begin{bmatrix} 0{,}3325 & 0{,}6675 \end{bmatrix}$$

O interpretare (partială) a acestui rezultat este: din cei 25% consumatori curenti ai produsului A, 88% vor continua acest obicei si din cei 75% dintre cumpărătorii produsului concurent, 15% vor comuta la produsul A. Asta dă un total (fractionar) de (0,25)(0,88) + (0,75)(0,15) = 0,3325 care vor cumpăra în perioada următoare produsul A.



Asadar, în anul 2, 33,25% din consumatori vor cumpăra produsul A, vor fi în starea 1. Ca verifcare, suma componentelor vectorului s_t trebuie să fie de fiecare dată egală cu unitatea.
Prin încă o multiplicare cu matricea P, starea sistemului în anul al treilea va fi

$$s_3 = s_2P = \begin{bmatrix}0,3325 & 0,6675\end{bmatrix}\begin{bmatrix}0,88 & 0,12\\ 0,15 & 0,85\end{bmatrix} = \begin{bmatrix}0,392725 & 0,607275\end{bmatrix}$$

Prin urmare, 32,2725% dintre consumatori vor cumpăra în anul trei produsul A.
Sub aspect practic este un nonsens a crede că se poate anticipa cota procentuală de piată pentru proudsul A la doi ani, cu patru zecimale. Dar evaluările permit o privire asupra perspectivei produsului A pe piată, perspectivă care altminteri nu este accesibilă.
Cu un program de calcul adecvat se pot face evaluări pentru perioade mai îndelungate. Iată mai sus graficul evolutiei cotei de piată a produsului A pe un interval de 25 de ani. Se observă o variatie rapidă în primii ani, o plafonare în anii următori anului 12.

Schimbări

Un avantaj al utilizării unui program este acela că efectele unei schimbări pot fi investigate usor. De exemplu, admitând că printr-o campanie de marketing/publicitate compania A poate creste loialitatea clientilor ei, în particular prin cresterea probabilitătii de a trece de la satrea A la starea A (de a rămâne fideli) cu 0,01, adică la 0,89.
Dacă compania A face acest lucru, e de asteptat ca si concurenta să-si lanseze propria campanie de promovare ceea ce aduce probabilitatea de a trece de la Concurentă la Concurentă de la 0,85 la, de pildă, 0,86.
După aceste evenimente, schimbări, compania A va avea o parte mai mare sau mai mică din piată (sau poate aceeasi). Prin calcul se poate anticipa rezultatul actiunii.
Calculul indică nu o crestere a vânzărilor în doi ani, ci o scădere: vechea sectiune de piată era 39,2725%, noua parte de piată este 38,5625%. A sti acest rezultat fără efortul si cheltuiala cu marketingul si cu publicitatea este un lucru foarte important.

Comportarea pe termen lung

Revenind la problema perspectivei pe termen lumg a produsului A, e necesar a calcula s_t pentru t foarte mare, a face o tentativă de a trece la limtă ($t \to \infty$). Ideea se bazează pe bănuiala că la un moment în viitor sistemul va atinge un echilibru, o stare stationară, în sensul că $s_t = s_{t-1}$. Faptul acesta nu înseamnă că tranzitiile între stări nu mai au loc. Ele se produc dar se produce o balansare, o

echilibrare asa încât numerele în s_t corespunzătoare fiecărei stări rămân aceleasi. Starea aceasta se numeste *starea stationară*.
Sunt două posibilităti de a sonda starea stationară: prin calcul direct sau apelând la algebră.
Calculul direct înseamnă repetarea evaluărilor pentru $t = 1, 2, 3, \ldots$ până când s_t si s_{t-1} sunt aproape la fel. Este, evident, calea cea mai facilă pentru calculator si este frecvent utilizată.
Calea algebrică evită calculele îndelungi. La starea stationară $s_t = s_{t-1}$ ($= [x_1\ x_2]$, de exemplu) si $s_t = s_{t-1} P$ si, în particular si mai în detaliu

$$[x_1 \quad x_2] = [x_1 \quad x_2]\begin{bmatrix} 0,88 & 0,12 \\ 0,15 & 0,85 \end{bmatrix}$$

cu $x_1 + x_2 = 1$. Asadar, aparent sunt de rezolvat trei ecuatii. Aceste trei ecuatii nu sunt totdeauna rezolvabile. De pildă sistemul pentru matricea de tranzitie

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

nu are o stare stationară.
Adoptând tratarea algebrică în cazul produsului A, sistemul este

$$x_1 = 0,88x_1 + 0,15x_2$$
$$x_2 = 0,12x_1 + 0,85x_2$$
$$x_1 + x_2 = 1$$

care după rearanjarea primelor două ecuatii se transformă în

$$0,12\mathrm{x}_1 - 0,15\mathrm{x}_2 = 0$$
$$0,12\mathrm{x}_1 - 0,15\mathrm{x}_2 = 0$$
$$\mathrm{x}_1 + \mathrm{x}_2 = 1$$

un sistem omogen si încă o ecuatie. Ecuatia ultimă este esentială. Fără ea nu se poate obtine o solutie unică. Prin rezolvare se obtin valorile $x_1 = 0,5556$ si $x_2 = 0,4444$, astfel că pe termen lung produsul A va avea o parte de piată de 55,56%.
O verificare numerică utilă – în particular pentru probleme de mai mari dimensiuni – constă în a substitui valorile finale calculate în ecuatiile initiale pentru a verifica compatibilitatea lor cu aceste ecuatii.

Comentarii

Analiza de mai sus este mai curând simplistă. Nu se poate pretinde că predictia portiunii de piată viitoare este foarte precisă. Schimbând circumstantele se schimbă în timp si matricea tranzitiilor. Cu toate acestea analiza de mai sus dă o oarecare anticipare asupra fenomenului inexistentă înaintea calculelor. De pildă:

- s-a creat o idee cantitativă asupra rapiditătii cu care partea de piată pentru produsul A se asteaptă a creste
- s-a creat o idee cantitativă asupra maximului asteptat al părtii de piată a produsului A

- s-a creat o idee cantitativă asupra efectului publicitătii asupra probabilitătilor de tranzitie si în ultimă instantă a efectului de crestere sau de scădere a părtii de piată pentru produsul A.

Asemenea evaluări bazate pe o analiză cantitativă poate fi de mare ajutor în procesele decizionale. De exemplu, dacă de doreste pentru A o sectiune de piată de 35% în doi ani, nu trebuie întreprins nimic dată fiind tendinta curentă. Dacă se doreste o proportie de 50% din piată în următorii doi ani atunci tot tinând seamă de tendintele curente trebuie întreprins ceva. Cât de mult trebuie schimbate probabilitătile de trecere pentru a atinge obiectivul de 50% din piată peste doi ani se calculează usor.

Aceste calcule sunt un suport de reflectie asupra efectului publicitătii. Pentru multe produse (deoarece cererea totală este efectiv stabilă, ceea ce se mai numeste uneori piată saturată, adică "toti oamenii care ies să cumpere cumpără") ceea ce face o campanie publicitară este a schimba probabilitătile de comutare (tranzitie) de la un produs la altul. De retinut că probabilitătile de tranzitie nu sunt numere fixe, ele pot avea propria lor evolutie.

Surse de date

În prezent, multe supermarketuri au introdus "carduri de fidelitate" care sunt prelucrate la iesire odată cu marcarea cumpărăturilor făcute de clienti. Acestea furnizează o cantitate de informatii detaliate, din care supermagazinele sau alte unităti economice interesate pot deduce matrici de tranzitie a preferintelor. Asemenea informatii sunt interesante si pentru
un producător de cereale pentru micul dejun. Cât ar costa un supermarket important colectarea continuă pe cale electronică de informatii detaliate privind comutarea între brand-urile de cereale? Cât ar putea încasa suplimentar pentru drepturi exclusive asupra informatiilor recoltate de acel supermagazin astfel ca concurenta să nu aibă acces la acele informatii? Cât de util este un flux permanent de astfel de informatii pentru a judeca efectul campaniilor de promotare/marketing? Sunt întrebări care sunt proprii economiei de piată.

Un supermarket vinde foarte multe produse diferite, de tipuri diferite. Datele pe care supermaketurile le colectează în bazele lor date prin cardurile de fidelitate pot fi extrem de valoroase pentru ele dar, poate mai ales, pentru producători. Accesul la asemenea date face posibilă construirea unor modele încă mai detaliate. De pildă, în problema cerealelor despre care s-a vorbit mai sus, concurenta a fost reprezentată printr-o singură (variabilă de) stare. Cu date mai detaliate acea stare poate fi descompusă într-un număr de stări diferite – poate una pentru fiecare brand de cereale în competitie pe piată. Dacă numărul de stări este n, numărul tranzitiilor este n^2 si sunt necesare tot atâtea probabilităti de tranzitie. Estimarea acestor probabilităti nu este o treabă foarte grea dacă se accesează o bază de date a unui supermarket. Din datele consumatorilor individuali se poate vedea dacă oamenii comută sau nu de la un produs la altul si dacă da, la ce anume comută.

Se pot imagina modele diferite pentru segmente diferite ale pietii. Poate că schimbarea dee brand-uri se face diferit în mediul rural si în mediul urban, de pildă. Familiile cu cu copii de vârste mici pot constitui un segment separat în spectrul de consumatori de cereale.
Este de retinut faptul că informatia cheie în investigarea numerică a comutării între brand-uri o reprezintă **probabilitătile asociate tranzitiilor**. Fără asemenea date nici un calcul nu este posibil.
Cum se pot obtine informatii relativ la probabilitătile de tranzitie dacă accesul, costisitor de cele mai multe ori, la informatiile adunate de supermarketuri nu este posibil? O cale este cunoscută încă dinainte de introducerea de carduri de fidelitate: supravegherea individuală a consumatorilor. Cineva se postează în iesirea supermarketului si chestionează cumpărătorii asupra cumpărăturilor curente si asupra cumpărăturilor precedente. Si calea aceasta poate costa destul de mult deoarece e necesar a acoperi o arie geografică suficientă si a face actualizări periodice ale acestor observatii.
Ambele căi, si colectarea electronică a datelor, si colectarea lor "manuală" costă bani.
Există o cale de a face supravegherea aceasta cu costuri mult reduse, cum se va vedea mai jos. Se iau în considerarea numai cotele de piată observate si putinătatea relativă a datelor se compensează cu o cantitate suplimentară de efort intelectual. Metoda realizează estimarea probabilitătilor de tranzitie – si a matricei de tranzitie – din împărtirea curentă, observată a pietei. Iată dezvoltarea imediat.

Estimarea matricei de tranzitie – două perioade

Iată un exemplu simplu cu două stări, referitoare la două companii. Se presupune că divizarea pietii în prima perioadă este [0,3 0,7] si în perioada următoare este [0,2 0,8]. Cum se estimează matricea de tranzitie?
Fie această matrice

$$\begin{bmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{bmatrix}$$

Se scrie ecuatia matricială

$$[0,2 \quad 0,8] = [0,3 \quad 0,7]\begin{bmatrix} p_1 & 1-p_1 \\ 1-p_2 & p_2 \end{bmatrix}$$

în necunoscutele p_1 si p_2 care se rescrie ca

$$-0,5 = 0,3p_1 - 0,7p_2$$
$$0,5 = -0,3p_1 + 0,7p_2$$

Se vede că cele două ecuatii sunt identice: una este cealaltă multiplicată cu –1, asadar solutia în p_1, p_2 nu este unică. Pentru a depăsi acest obstacol se tine seamă de faptul că oricare ar fi matricea de tranzitie, elementele diagonale (p_1 si p_2) ar trebui să fie cât mai mari posibile. Asta este echivalent cu a spune că este mai credibilă constanta optiunilor si mai putin probabilă comutarea la alt brand. Asadar, se poate formula o problemă de forma:

$$\text{A se maximiza } p_1 + p_2$$

în conditiile

$$-0{,}3p_1 + 0{,}7p_2 = 0{,}5$$
$$0 \le p_1 \le 1$$
$$0 \le p_2 \le 1$$

cu includerea evidentă a restrictiilor uzuale asupra probabilitătilor, care trebuie să fie numere subunitare si pozitive. Această problemă, în formularea nouă este una de programare liniară usor de rezolvat.
Solutia este $p_1 = 2/3$ si $p_2 = 1$, adică

$$P = \begin{bmatrix} 2/3 & 1/3 \\ 0 & 1 \end{bmatrix}$$

Este usor de verificat că valorile stabilite satisfac ecuatia matricială de mai devreme. De observat că există multe matrici de tranzitie posibile care corespund exact cotelor de piată observate. Pe calea arătată s-a ales una din acele matrici de tranzitie, poate nu cea mai potrivită.

Estimarea matricei de tranzitie – perioade multiple

Cercetarea capătă consistentă dacă datele culese din realitate sunt mai bogate, cum ar fi de pildă în cazul de mai sus dacă se presupune că observatiile asupra cotelor de piată pe perioade egale sunt: [0,3 0,7], apoi [0,2 0,8], apoi [0,15 0,85], apoi [0,13 0,87]. Primele două perioade sunt cele utilizate în evaluările de mai devreme. Cum se poate estima matricea de tranzitie în noile conditii?
O cale imediată ar fi să se ia perechi succesive de vectori din secventa de mai sus si să fie tratate prin metoda deja expusă. Aproape sigur, matricile de tranzitie vor rezulta diferite si se vor aplica pentru fiecare pereche de vectori ai cotelor de piată în timpul une perioade. Stabilirea acestor matrici ar putea fi un exercitiu pentru cititor.
Această variatie a matricei de tranzitie de la o perioadă la alta scoate discutia din aria solidei teorii a lanturilor Markov.
Gândul ar putea duce la formularea unei probleme de programare liniară mai bogată în conditii restrictive: la relatia liniară din formularea de mai sus s-ar putea adăuga alte relatii liniare generate de perechile următoare de vectori ai cotelor de piată. Problema ar fi atunci:

$$\text{A se maximiza } p_1 + p_2$$

în conditiile

$$0{,}3p_1 + 0{,}7(1 - p_2) = 0{,}2$$
$$0{,}3(1 - p_1) + 0{,}7p_2 = 0{,}8$$
$$0{,}2p_1 + 0{,}8(1 - p_2) = 0{,}15$$
$$0{,}2(1 - p_1) + 0{,}8\mathrm{p}_2 = 0{,}85$$
$$0{,}15p_1 + 0{,}85(1 - p_2) = 0{,}13$$
$$0{,}15(1 - p_1) + 0{,}85\, p_2 = 0{,}87$$

$$0 \le p_1 \le 1$$
$$0 \le p_2 \le 1$$

Evident, restrictiile-egalitate coincid două câte două, asadar sunt numai trei egalităti distincte.

Încercarea de rezolvare prin metodele programării liniare este un esec: problema este infezabilă, adică nu există valori p_1 si p_2 care să satisfacă conditiile de mai sus.

Există o altă cale de atac, care se constituie tot ca o problemă de optimizare: minimizarea diferentelor la pătrat dintre probabilitătile stărilor calculate pentru fiecare perioadă si stările corespunzătoare observate, adică minimizarea sumei de pătrate ale erorilor de predictie a cotelor de piată estimate cu modelul Markov. Aceasta este o tratare uzuală în estimarea de parametri.

Asadar, este de rezolvat prolema:

A se minimiza functia de p_1 si p_2

$$[0,3\,p_1 + 0,7(1-p_2) - 0,2]^2 + [0,3(1-p_1) + 0,7\,p_2 - 0,8]^2 +$$
$$+ [0,2\,p_1 + 0,8(1-p_2) - 0,15]^2 + [0,2(1-p_1) + 0,8\,p_2 - 0,85]^2 +$$
$$+ [0,15\,p_1 + 0,85(1-p_2) - 0,13]^2 + [0,15(1-p_1) + 0,85\,p_2 - 0,87]^2$$

în conditiile

$$0 \le p_1 \le 1$$
$$0 \le p_2 \le 1$$

Putin calcul algebric si eliminarea unor constante care nu influentează pozitia minimului în spatiul (p_1, p_2) aduc functia obiectiv de optimizat la forma de mai jos si problema la a se minimiza

$$0,1525(p_1)^2 + 1,8525(p_2)^2 - 0,995p_1p_2 + 0,776p_1 - 2,964p_2$$

în conditiile mentionate.

Este vorba aici de o problemă de programare neliniară (pătratică) care se rezolvă cu un program specializat.

Rezultatul este $p_1 = 0,53$, $p_2 = 0,94$ si, în consecintă, matricea de tranzitie este

$$P = \begin{bmatrix} 0,53 & 0,47 \\ 0,06 & 0,94 \end{bmatrix}$$

Cu aceste valori se pot calcula stările succesive.

Timpul	Starea	Estimat	Observat
1	1	0,201	0,2
	2	0,799	0,8
2	1	0,145	0,15
	2	0,855	0,85
3	1	0,131	0,13
	2	0,869	0,87

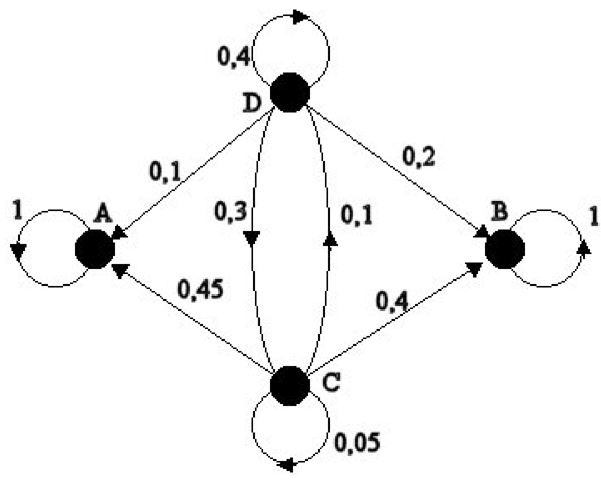
În tabel se remarcă o estimare foarte bună prin matricea de tranzitie a cotelor de piată observate efectiv.

Pentru claritatea lucrurilor: ceea ce s-a lucrat până la acest punct a avut ca urmare găsirea într-o manieră sistematică, consistentă logic, a unei matrici de

tranzitie care se potriveste cel mai bine pe cotele de piată observate. Acea matrice de tranzitie poate să corespundă sau nu probabilitătilor de tranzitie pe care le-am fi aflat prin monitorizarea clientilor sau prin colectarea electronică a informatiei din lumea reală.
Totusi, matricea de tranzitie stabilită mai sus dă o imagine întrucâtva mai completă asupra situatiei. Se poate observa fidelitatea clientilor companiei 2 (94% rămân cu compania 2, în starea 2, perioade la rând). Clientii companiei 1 (starea 1) nu sunt la fel de loiali: numai 53% rămân cu compania 1 perioadă după perioadă, aproape echivalent cu a arunca cu banul pentru a alege între produsele companiei 1 si ale companiei 2. Această apreciere numerică a fidelitătii este ceva ce nu putea fi făcut privind pur si simplu la valorile cotelor de piată observate: [0,3 0,7], apoi [0,2 0,8], apoi [0,15 0,85], apoi [0,13 0,87]. Aceeasi evaluare sugerează căile spre câstigarea unei cote de piată ma ridicată sau de stopare a declinului eventual al cotei de piată.

Un exemplu mai complicat

Pentru consolidarea cunostintelor despre procesele Markov se propune exemplul următor. Se admite că piata pentru un anumit produs este alimentată/controlată de patru companii: A, B, C si D. Dacă clientii cumpără produse de tipul A sau B ei nu mai cumpără vreodată alt brand. Dacă cumpără C probabilitătile de a cumpăra luna viitoare A, B, C sau D sunt 0,45; 0,4; 0,05 respectiv 0,1. Dacă ei cumpără D probabilitătile ca luna viitoare să cumpere A, B, C sau D sunt 0,1; 0,2; 0,3 respectiv 0,4.
Situatia aceasta se prezintă ca în diagrama de tranzitie a stărilor dată mai sus.
Dacă cumpărătorii sunt distribuiti initial în proportia 20%, 30%, 30% si 20% pentru A, B, C respectiv D, care va fi situatia după două luni?



Punând starea 1 = A, starea 2 = B, starea 3 = C si starea 4 = D, se poate scrie

$$s_1 = [0{,}2 \quad 0{,}3 \quad 0{,}3 \quad 0{,}2]$$

si

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0{,}45 & 0{,}4 & 0{,}05 & 0{,}1 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}5 \end{bmatrix}$$

Stările 1 si 2 (A si B) sunt stări *absorbante*, sunt stări care odată atinse nu mai pot fi părăsite. Stările care nu sunt absorbante sunt numite stări *tranzitorii*.
Starea sistemului în luna a doua este dată de $s_2 = s_1P$

$$[0{,}2 \quad 0{,}3 \quad 0{,}3 \quad 0{,}2]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0{,}45 & 0{,}4 & 0{,}05 & 0{,}1 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}5 \end{bmatrix} = [0{,}355 \quad 0{,}46 \quad 0{,}075 \quad 0{,}11]$$

În luna a treia starea se se modifică la $s_3 = s_2P$

$$[0{,}355 \quad 0{,}46 \quad 0{,}075 \quad 0{,}11]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0{,}45 & 0{,}4 & 0{,}05 & 0{,}1 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}5 \end{bmatrix} =$$

$$= [0{,}39975 \quad 0{,}512 \quad 0{,}03675 \quad 0{,}0515]$$

Se observă că pentru produsele A si B cotele de piată sunt în crestere, cotele de piată pentru produsele C si D sunt în scădere de la perioadă la perioadă.
Repetarea calcului până la perioada a 20-a produce un rezultat asteptat: cumpărătorii migrează treptat la produsele A si B. Starea după 20 de perioade este

$$[0{,}4407 \quad 0{,}5593 \quad 4{,}815.10^{-8} \quad 6{,}761.10^{-8}]$$

Practic, toti consumatorii se adună, în cele din urmă, în stările absorbante.
De observat că existenta unor stări absorbante face matematica necesară calculului evolutiei sistemului pe durată îndelungată mult mai complicată decât aceea utilizată mai devreme, în cazurile în care nu existau stări absorbante. Programul de calcul obisnuit refuză să calculeze în acest caz un regim stationar. Pentru a vedea de ce, se încearcă aceeasi metodă ca mai devreme, utilizată în lipsa stărilor absorbante.
Fie starea finală a sistemului $[x_1 \ x_2 \ x_3 \ x_4]$. Atunci se caută solutia ecuatiei

$$[x_1 \quad x_2 \quad x_3 \quad x_4] = [x_1 \quad x_2 \quad x_3 \quad x_4]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0{,}45 & 0{,}4 & 0{,}05 & 0{,}1 \\ 0{,}1 & 0{,}2 & 0{,}4 & 0{,}5 \end{bmatrix}$$

cu $x_1 + x_2 + x_3 + x_4 = 1$.
Ultimele două ecuatii din ecuatia matricială sunt

$$x_3 = 0{,}05x_3 + 0{,}3x_4$$
$$x_4 = 0{,}1x_3 + 0{,}4x_4$$

ceea ce conduce la $x_3 = x_4 = 0$, singurele valori care verifică cele două egalităti (o bănuială în acest sens exista). Cu $x_3 = x_4 = 0$, ecuatiile care rămân devin

$$x_1 = x_1$$
$$x_2 = x_2$$
$$x_1 + x_2 = 1$$

ccea ce nu duce nicăieri.

Orice problemă pentru care se poate desena o diagramă de tranzitie a stării în genul figurat mai devreme poate fi analizată prin metoda dată mai sus.

Avantajele si dezavantajele utilizării teoriei lui Marcov sunt:

- Teoria markoviană este simplu de înteles si de aplicat
- Calculele de sensibilitate (problemele de genul "dar dacă") sunt usor de efectuat
- Teoria lui Markov dă o privire asupra evolutiei sistemului în timp
- Matricea P poate fi dependentă de starea curentă a sistemului. Dacă P depinde atât de timp cât si de starea curentă a sistemului, adică P este o functie de t si de s_t atunci ecuatia Markov de bază se complică. Ea devine $s_t = s_{t-1}P(t-1, s_{t-1})$
- Teoria lui Markov este numai un model simplificat al proceselor decizionale reale.

O aplicatie interesantă a proceselor Markov comentată în literatură se referă la industria petrolieră off-shore norvegiană. În Norvegia, un organism de stat, The the Norwegian Petroleum Directorate, împreună cu compania petrolieră de stat STATOIL are un rol important în planificarea dezvoltării facilitătilor petrol-gaze off-shore.

Problema principală si esentială pe care o are The Norwegian Petroleum Directorate este cum să planifice conductele, pornirile din teren, productia astfel încât să maximizeze contributia la economia norvegiană în timp. Scara de timp este aici foarte îndelungă, tipic de la 30 la 50 de ani.

Există flexibilitate în deciziile relative la un număr de aspecte cum sunt:

- Rata productiei (cât de repede iese produsul din subsol)
- Initierea de noi exploatări (când ar trebui pornite exploatările)
- Constructia si capacitatea conductelor (unde să fie construite, când să fie construite si de ce capacitate ar trebui să fie.

Obiectivul este acela de a maximiza beneficiul economiei norvegiene în timp, peste ani.

De importantă critică este pretul titeiului – desi nu poate fi prezis cu acuratete pe perioade lungi, de 30 la 50 de ani.

Pentru a depăsi această problemă norvegienii modelează pretul petrolului ca un proces Markov cu trei niveluri (stări), care corespund unor scenarii: unul optimist, unul care pare cel mai probabil si altul pesimist. Ei specifică totodată probabilităti asociate tranzitiilor între stări pentru fiecare perioadă de timp (an). Se pot utiliza matrici de tranzitie diferite pentru scări de timp diferite (de pildă

matrici diferite pentru viitorul apropiat, pe termen mediu si pe perspectivă îndepărtată).
Desi tratarea este destul de simplă, ea prezintă avantajul captării incertitudinii viitorului într-un model relativ simplu, usor de înteles si de aplicat.
Studiile de modelare a populatiilor (cu obiecte care "îmbătrânesc") sunt si ele aplicatii interesante ale proceselor Markov. Un exemplu de gen este modelarea pietii automobilelor ca proces Markov, în vederea prognozării "necesarului" de automobile noi pe măsură ce vehiculele vechi ies din uz. Pentru a vedea asta se poate încerca modelarea pietei de gen cu stări corespunzătoare numărului de proprietari/vechimea vehiculelor.
Un alt exemplu este modelarea ca proces Markov a evolutiei clinice a unui pacient sub tratament cu diferite medicamente.

GRAFURI SI APLICATII ALE GRAFURILOR

Generalităti si definitii

Grafurile sunt obiecte definite matematic ca perechi (X, Γ) cu X o multime si Γ o aplicatie definită pe multimea X cu valori în multimea părtilor lui X. Dacă X este o multime finită atunci unui graf i se poate asocia o reprezentare geometrică prin puncte si segmente. Punctele, numite si noduri sunt elemente ale lui X, iar segmentele, numite si arce exprimă aplicatia Γ. Un arc (orientat) are ca origine un element din X si ca extremitate un nod din submultimea imagine prin Γ a nodului de origine, parte a multimii X. Un graf poate fi definit si prin cuplul (X, U) cu U multimea arcelor.
Structura grafurilor poate fi foarte diferită. Ele pot fi *conexe* sau *neconexe* după cum există sau nu un drum între oricare două noduri ale grafului. *Drum* este orice succesiune de arce cu extremitatea si/sau originea coincidente. Grafurile pot fi *orientate* sau nu după cum sensul parcurgerii arcelor este important sau nu. Grafurile pot fi *ciclice* sau nu după cum există sau nu un drum parcurs în sensul orientării arcelor, care să pornească într-un nod si să revină în acel nod.
Frecvent, arcelor unui graf li se asociază anumite numere care sunt cunoscute generic drept *capacitătile* acelor arce.
Grafurile au aplicatii multiple în modelarea si simularea sistemelor economice. Câteva din aceste aplicatii sunt discutate în continuare.

Analiza drumului critic

Analiza drumului critic (ADC) este o aplicatie economică remarcabilă a grafurilor. Aceasta este o metodă de conducere stiintifică a realizării proiectelor. Un proiect este un proces complex sau o actiune de mare amploare orientată spre atingerea unui scop bine precizat. Un proiect, în sensul acestei definitii are un obiectiv, un ansamblu de activităti si o tehnologie.
Activitătile sunt părti determinate ale proiectului, care consumă uzual timp si resurse. Descompunerea unui proiect în activităti componente permite analiza amănuntită a desfăsurării lui în conformitate cu tehnologia pe care el se bazează.
A programa un proiect înseamnă a stabili termenele de începere pentru fiecare activitate tinând seama, din nou, de logica tehnologiei proiectului. Din multitudinea programelor admisibile este de retinut un număr restrâns de programe optime, uneori unul singur. Un astfel de program asigură

optimizarea unui anumit criteriu de eficientă economică fără a viola conditiile tehnologice.
Ordinea si conditionarea tehnologică a activitătilor unui proiect se poate modela prin grafuri în cel putin două feluri după modul în care se plasează activitătile componente ale proiectului în graful-model. Sunt grafuri cu activitătile în noduri sau grafuri cu activitătile pe arce.
În oricare din cele două variante, începutul constă în întocmirea unei liste a activitătilor care compun proiectul si cu stabilirea precedentelor. Operatia implică desigur o gândire care trebuie să tină seamă de tehnologia realizării proiectului.
Ca *exemplu*, se dă imediat lista de activităti în cazul unui proiect simplu, alcătuit din 11 activităti cu duratele lor de executie trecute în tabelul alăturat. Se presupune aici că trebuie făcută o reproiectare (minoră) a unui produs si a ambalajului său. Intentia este a verifica piata pentru acest produs reproiectat si apoi a-l revizui în raport cu rezultatele testului de piată. În final, concluziile sunt prezentate în fata conducerii (colective a) companiei.
Întrebarea cheie este: cât timp va consuma acest proiect cu întelesul cât de redus poate fi acest timp?

Numărul activitătii	Scurtă descriere	Durata
1	Reproiectarea produsului	6
2	Reproiectarea ambalajului	2
3	Comandarea si primirea componentelor pentru produsul reproiectat	3
4	Comandarea si primirea componentelor pentru ambalajul reproiectat	2
5	Asamblarea produselor	4
6	Pregătirea ambalajului	1
7	Împachetarea produsului reproiectat	1
8	Testarea pe piată a produsului reproiectat	6
9	Revederea produsului reproiectat	3
10	Revederea ambalajului reproiectat	1
11	Prezentarea rezultatelor în fata conducerii	1

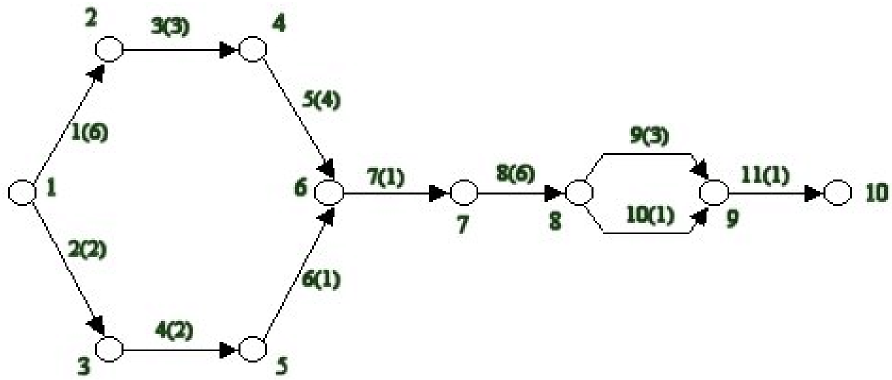
Relativ la alcătuirea acestei liste de activităti se poate totdeauna pune în discutie gradul de detaliere a proiectului în timp (scara de timp). La o extremă se putea considera proiectul ca o activitate unică, "executarea proiectului"; la cealaltă extremă proiectul se poate fărâmita în activităti pe ore. Scara de timp adecvată, care poate fi uneori diferită pentru activităti diferite rezultă din cunoasterea situatiei concrete combinată cu experienta.
Alături de această listă trebuie pregătită o listă secundară, dar nu mai putin importantă, cu relatiile de precedentă, indicatoare ale logicii succesiunii activitătilor. Această listă spune care activităti trebuie încheiate înainte ca alte activităti să poată începe. De pildă, în lista de mai sus, înainte ca activitatea 3 să poată începe, trebuie finalizată activitatea 1. Pentru claritate, această listă ar trebui mentinută la un minim de informatie, prin specificarea numai a relatiilor imediate, adică numai a relatiilor care implică activităti

adiacente în timp. De exemplu, este evident că activitatea 1 trebuie finalizată înainte de începerea activitatătii 9, dar despre aceste două activităti cu greu se poate spune că au o relatie imediată, deoarece multe alte activităti următoare activitătii 1 trebuie să fie încheiate înainte de startul activitatătii 9. În schimb, activitătile 8 si 9 sunt activităti care au o relatie imediată: activitatea 9 poate începe deîndată ce activitatea 8 este încheiată. Prin specificarea relatiilor care nu sunt imediate, lucrurile mai curând se complică, la fel calculele de executat, fără a afecta însă rezultatul final. Nu-i mai putin adevărat că în raport cu lumea reală, consecintele omiterii unor relatii de precedentă sunt mult mai serioase decât consecintele includerii unor relatii nenecesare, care nu sunt imediate.
Iată acum lista precedentelor imediate pentru activitătile componente ale proiectului simplu exemplificat:

Activităti precedente	Activitatea următoare
1	3
2	4
3	5
4	6
5, 6	7
7	8
8	9
8	10
9, 10	11

De observat că:

- Activitătile 1 si 2 nu apar în coloana din dreapta a tabelului de precedente deoarece nu există activităti care trebuie încheiate înainte ca ele să poată fi începute. Activitătile 1 si 2 pot fi pornite imediat
- Două activităti (5 si 6) trebuie finalizate înainte ca activitatea 7 să poată începe
- Este destul de clar în acest tabel că relatiile de precedentă neimediate (de genul activitatea 1 trebuie încheiată înainte ca activitatea 9 să poată fi începută) nu trebuie incluse în listă deoarece ele pot fi deduse din relatiile deja prezente în listă.



În această fază, există informatii suficiente pentru a construi un graf-model al proiectului.
În varianta cu activitătile pe arce acesta arată ca în figura alăturată.
Graful din figură este cunoscut în literatură ca un graf-retea sau, simplu, o retea CPM (de la *Critical Path Method*). Nodurile numerotate distinct marchează evenimente care constau în încheierea unor activităti si crearea posibilitătii de începere a altora. Exceptii fac nodul de început (1 în exemplul în discutie) si nodul final (10). Ca regulă generală atât nodul de început cât si cel care marchează finalul proiectului sunt unice. După cum se va vedea în altă sectiune a acestei lucrări, activitătile pot avea durate aleatoare. Nodurile sunt asociate atunci unor evenimente în sensul discutat în capitolul de probabilităti. Este cazul retelelor PERT (de la *Program Evaluation and Review Technique*).
În cazul retelelor CPM duratele activitătilor sunt determinate. Nodurile (evenimentele) se nomerotează, cum s-a spus, cu numere naturale 1, 2, …, *i*, …, *j*, …, *n*, de preferintă fără lacune. O pereche de astfel de numere, (*i*, *j*) marchează o activitate cu începutul în nodul *i* si cu sfârsitul în nodul *j*. Activitătile si termenele lor se reprezintă dacă este posibil pe graf. În rationamentele care urmează duratele si termenele se notează astfel:
t_{ij} - durata activitătii (*i*, *j*);
t_i - termenul cel mai timpuriu (minim) al unui eveniment/nod (*i*);
t_i^* - termenul cel mai târziu (maxim) al unui eveniment/nod (*i*);
$t_s(i,j)$ - termenul minim de începere a activitătii (*i*, *j*);
$t_s^*(i,j)$ - termenul maxim de începere a activitătii (*i*, *j*);
$t_f(i,j)$ - termenul minim de încheiere a activitătii (*i*, *j*);
$t_f^*(i,j)$ - termenul maxim de încheiere a activitătii (*i*, *j*);
T - durata totală a proiectului;
$R_t(i,j)$ - rezerva totală a activitătii (*i*, *j*);
$R_l(i,j)$ - rezerva liberă a activitătii (*i*, *j*);
$R_i(i,j)$ - rezerva intermediară a activitătii (*i*, *j*);
$R_s(i,j)$ - rezerva sigură a activitătii (*i*, *j*);
Termenele cel mai timpuriu posibil si cel mai târziu admis pentru o activitate (*i*, *j*) se calculează din termenele evenimentelor care marchează începutul si sfârsitul ei, cu relatiile

$$t_s(i,j) = t_i$$
$$t_f(i,j) = t_i + t_{ij}$$
$$t_f^*(i,j) = t_j^*$$
$$t_s^*(i,j) = t_j^* - t_{ij}$$

Rezervele de timp de cele patru tipuri ale unei activităti (*i*, *j*) se obtin din termenele evenimentelor cu relatiile

$$R_t(i,j) = t_j^* - t_i - t_{ij}$$
$$R_l(i,j) = t_j - t_i - t_{ij}$$

$$R_i(i,j) = t_j^* - t_i^* - t_{ij}$$
$$R_s(i,j) = t_j - t_i^* - t_{ij}$$

Acestea sunt rezerve care pot fi consumate în amumite conditii fără a afecta durata totală de executie a proiectului. Din secventa de relatii de mai sus pot rezulta uneori valori negative. Desigur astfel de valori nu au sens practic si de aceea se consideră a fi semnul inexitentei acelor rezerve, nulitatea lor.

În practică, foarte frecvent se utilizează drept criteriu de optimizare durata totală a proiectului, care trebuie să fie, se întelege, cât mai scurtă. Algoritmul de rezolvare a problemei în acest caz are două etape. În prima etapă, aceea a *parcursului direct*, se calculează termenele minime ale evenimentelor, iar în etapa a doua, cea a *parcursului invers*, se calculează termenele maxime ale evenimentelor. Formulele de calcul sunt

$$t_j = \begin{cases} 0; & j = 1 \\ \max(t_i + t_{ij});\ 1 < j \le n \\ (i,j) \in G \end{cases}$$

respectiv

$$t_i^* = \begin{cases} t_n; & i = n \\ \min(t_j^* - t_{ij});1 \le i < n \\ (i,j) \in G \end{cases}$$

Parcursul direct, de la nodul initial spre cel final reprezintă un program minorant de executie a proiectului. Parcursul invers, de la nodul final spre cel initial este un program majorant. Încadrarea termenelor asociate nodurilor (evenimentelor) între cele două programe nu modifică termenul final si durata totală a proiectului. Nodurile pentru care

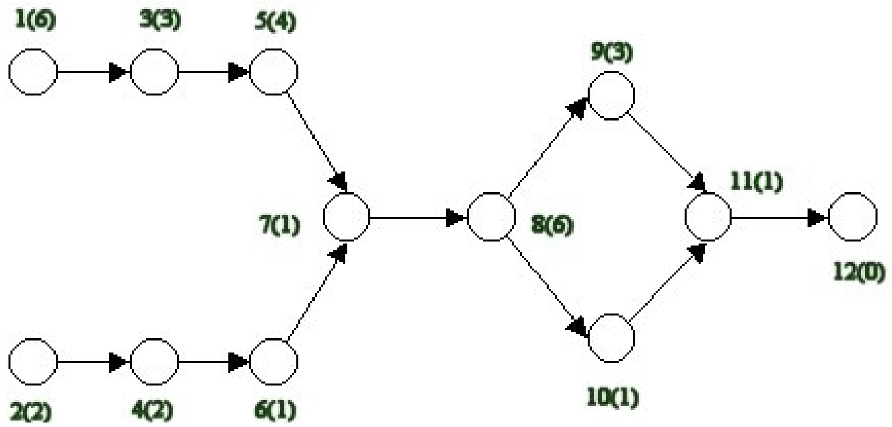
$$t_i = t_i^*$$

sunt noduri sau evenimente *critice*. Acestea sunt situate pe asa-numitul *drumul critic* si termenul unic, minim si maxim trebuie respectat riguros pentru fiecare nod. Toate celelalte evenimente admit o întârziere maximă de $t_i^* - t_i$. Acestea nu sunt situate pe drumul critic, sunt necritice.

Activitătile critice sunt situate între noduri critice si în cursul executării lucrărilor proiectului trebuie supravegheate îndeaproape deoarece orice prelungire a duratei unei activităti critice produce o întârziere a finalizării proiectului.

Despre rezervele de timp ale activitătilor trebuie spus că gestionarea lor trebuie făcută cu prudentă. Epuizarea lor poate produce criticalizarea unor activităti următoare, ceea ce, evident, complică managementul proiectului în continuare. Singurele rezerve care pot fi consumate fără a modifica numărul de activităti critice sunt rezervele sigure, adică cele din ultima categorie.

Graful cu activitătile în noduri pentru problema enuntată mai devreme arată ca în figura următoare. Graful în această variantă dă expresie grafică mai clară relatiilor de precedentă. Sunt noduri făre vreun precedent, care corespund activitătilor de început în desfăsurarea lucrărilor proiectului. Uzual, se mai introduce un nod (aici nodul 12) care este “activitatea” finală care marchează finalul punerii în operă a proiectului si nu consumă timp (nodurile sunt numerotate cu numerele asociate activitătilor si între paranteze sunt înscrise duratele acelor activităti).



Stabilirea termenelor celor mai timpurii posibile si a celor mai târzii admise pentru fiecare activitate urmează acelasi algoritm prezentat mai devreme, derivat din programarea dinamică.
Rezultatele obtinute (de regulă pe calculator) au forma din tabelul următor:

Activităti	Durate	Startul cel mai timpuriu	Finalul cel mai timpuriu	Startul cel mai târziu	Finalul cel mai târziu	Rezerve
1*	6	0	6	0	6	0
2	2	0	2	8	10	8
3*	3	6	9	6	9	0
4	2	2	4	10	12	8
5*	4	9	13	9	13	0
6	1	4	5	12	13	8
7*	1	13	14	13	14	0
8*	6	14	20	14	20	0
9*	3	20	23	20	23	0
10	1	20	21	22	23	2
11*	1	23	24	23	24	0
Durata proiectului = 24 u.t.						
Numărul de drumuri critice = 1 (caracterul * marchează activitătile critice)						

Tabelul rezultatelor evidentiază lipsa oricăror rezerve de timp pentru activitătile situate pe drumul critic. Apar în schimb rezerve pentru activiutătile necritice. Aceste rezerve sunt gestionabile optim dacă conducătorul lucrărilor de punere în operă a proiectului are în vedere, de pildă, o nivelare a resurselor necesare realizării proiectului. Acest Subiect va fi discutat într-o altă sectiune a acestei lucrări.

Lungimea drumului critic – suma duratelor ectivitătilor critice – este durata minimă a executării proiectului în conditiile din enunt.
Pot exista mai multe drumuri critice, dar toate au aceeasi lungime, aceeasi durată.

Metoda de analiză a drumului critic CPM cu reducere de durate.

În analiza drumului critic prin metoda CPM este posibil ca durata de executarea a proiectului rezultată să fie neconvenabil de lungă. Este posibilă accelerarea executiei unui proiect? Răspunsul este afirmativ dar, cum este de asteptat, costurile de executie cresc.
Dacă termenul de încheiere a proiectului trebuie redus, o seamă de activităti urmează a se executa în termene mai strânse. Desigur, candidatele prime la accelerare sunt activitătile de pe drumul critic: activităti mai scurte pe drumul critic înseamnă o durată însumată mai mică.
Scurtarea duratelor pe drumul critic ar putea să facă critice alte activităti, mai întâi dintre cele imediat adiacente drumului critic stabilit pentru durate normale dar nu numai. În faze mai avnsate ale evaluărilor poate deveni necesară scurtarea si a altor ectivităti. Din acest motiv, în general trebuie analizat si stabilit în prealabil costul scurtării oricărei activităti critică sau necritică din proiect si reanalizată programarea lucrărilor pentru a cheltui cât mai putin cu accelerarea necesară. Pentru costul reducerii duratei unei activităti, altfel oarecare, modelul cel mai frecvent utilizat este cel cu variatie liniară: se stabileste o durată limită sub care activitatea respectivă nu poate fi executată, $t_{ij\min}$ si se exprimă costul activitătii la durate intermediare, mai mari decât $t_{ij\min}$, mai mici în raport cu durata normală $t_{ij\text{normal}}$. Relatia cost-durată nu poate fi decât cu pantă negativă

$$C_{ij\,\text{redus}} = C_{ij\,\text{normal}} + c_{ij}\,(t_{ij\,\text{normal}} - t) \text{ pentru } t \in [t_{ij\,\min}, t_{ij\,\text{normal}}]$$

Minimizarea costului reducerii duratei de executie a proiectului parcurge un algoritm iterativ în care se alternează programarea liniară cu analiza drumului critic. La început se reduce durata acelei activităti de pe drumul critic, care are cea mai mică variatie c_{ij}, până când rezerva de reducere a duratei acelei activităti este epuizată sau până când drumul critic se modifică structural prin includerea altor activităti. Se reevaluează drumul critic dacă este cazul si se reia calculul cu o altă activitate de pe drumul critic (nou), care are rezerve de reducere a duratei si are cel mai mic cost specific c_{ij}. Se opreste calculul fie atunci când o conditie de durată este îndeplinită, fie când nu mai sunt posibile reduceri de durată.
Iată, ca exemplu, elementele de calcul necesare unui astfel de calcul, sub formă de tabel (cu referire la graful cu activitătile pe arce).

Activitatea	Noduri		Durate		Costuri	
	De început	De final	Normale	Scurtate	Normale	Cu scurtare
1	1	2	6	4	6000	8000
2	1	3	2	1	2000	5000
3	2	4	3	1	3000	7000
4	3	5	2	1	2000	4000
5	4	6	4	2	4000	9000
6	5	6	1	1	1000	1000
7	6	7	1	1	1000	1000
8	7	8	6	4	6000	7500
9	8	9	3	2	3000	4500
10	8	9	1	1	1000	1000
11	9	10	1	1	1000	1000

Mai devreme s-a văzut că durata normală a executiei proiectului este 24 de unităti de timp (u.t.). Se cere o scurtare cu 4 unităti de timp, adică la numai 20 u.t. Rezultatele sunt cuprinse în tabelul imediat următor.

Activitatea	Noduri		Scurtare cu:	Costul scurtării	Durata activitătii	Costul activitătii
	De început	De final				
1*	1	2	2	2000	4	8000
2	1	3	-	-	2	5000
3*	2	4	-	-	3	7000
4	3	5	-	-	2	4000
5*	4	6	-	-	4	9000
6	5	6	-	-	1	1000
7*	6	7	-	-	1	1000
8*	7	8	2	1500	4	7500
9*	8	9	-	-	3	3000
10	8	9	-	-	1	1000
11*	9	10	-	-	1	1000
Costuri:				3500		33500
Durata normală a proiectului: 24 u.t.						
Durata scurtată a proiectului: 20 u.t.						

Rezultatele din tabel contine politica de scurtare a duratei de la 24 la 20 de u.t. cea mai putin costisitoare.

Problema drumului critic în conditii de incertitudine

Asa cum s-a mentionat în treacăt într-o sectiune a capitolului prezent, uneori, asupra duratelor de executie a activitătilor dintr-un proiect planează incertitudinea. Nu se pot preciza valori sigure ci numai estimări de încadrare: durata unei activitătii (*i*, *j*) poate avea orice valoare dintr-un interval compact delimitat de un minim care este asociat celor mai fericite conditii de executie si de un maxim asociat cu cele mai adverse conditii în timpul executiei. Asemenea situatii se întâlnesc, de exemplu, atunci când una sau mai multe activităti sunt de noutate absolută si lipsa experientei exclude orice posibilitate de normare clară sau când asupra ritmului lucrărilor pentru ducerea la bun sfârsit a unor activităti impietează factori aleatori naturali sau care apartin de ambianta economică.

Pentru modelarea unor proiecte sub incertitudini de acest gen se utilizează grafuri de genul celor descrise mai sus. Metoda de tratare este diferită si este cunoscută sub numele prescurtat PERT (*Program Evaluation and Review Technique*).

Lucrurile stau întrucâtva diferit în cazul retelelor PERT unde duratele activitătilor sunt incerte. Pentru fiecare activitate din proiect, se estimează pe o cale sau alta o durată optimistă a_{ij}, o durată pesimistă b_{ij} si o durată care pare a fi cea mai probabilă m_{ij}. Cu aceste estimări primare, în ideia că durata unei activităti este o variabilă aleatoare, se pot evalua mediile si dispersiile duratelor fiecărei activităti, uzând de relatiile simplificate si aproximative

$$\bar{t}_{ij} = \frac{a_{ij} + 4m_{ij} + b_{ij}}{6}$$

$$\sigma_{ij}^2 = \left(\frac{b_{ij} - a_{ij}}{6} \right)^2$$

Legea de repartitie cea mai potrivită pentru duratele activitătilor într-o retea PERT este o lege Beta cu densitatea de repartitie sau, cum i se mai spune, densitatea de probabilitate

$$f(t) = \begin{cases} 0 & t < a \\ \dfrac{(t-a)^p (b-t)^q}{(b-a)^{p+q+1} B(p+1, q+1)} & a \le t \le b \\ 0 & t > b \end{cases}$$

Legea aceasta contine doi parametri, p si q, numere pozitive, iar $B(p + 1, q + 1)$ este functia specială euleriană de specia a doua definită ca

$$B(p+1, q+1) = \int_0^1 x^p (1-x)^q \, dx$$

Variabila aleatoare t are media

$$\bar{t} = \frac{a + b + (p+q)m}{p+q+2} \quad \text{în care} \quad m = \frac{aq + bp}{p+q}$$

si dispersia

$$\sigma_t^2 = \frac{(b-a)^2 (p+1)(q+1)}{(p+q+3)(p+q+2)^2}$$

Relatiile estimative date mai sus în functie de cele trei valori a_{ij}, b_{ij} si m_{ij} sunt aproximări ale acestor valori exacte în expresia cărora constantele a, b, m sunt înlocuite a_{ij}, b_{ij}, respectiv cu m_{ij}, iar exponentii p si q nu sunt foarte diferiti.

Algoritmul pentru stabilirea drumului critic este acelasi ca si în cazul determinist, numai că se folosesc duratele medii ale activitătilor evaluate aproximativ sau exact conform relatiilor deja prezentate. Desigur, durata proiectului este o variabilă aleatoare. Media ei este suma duratelor medii ale activitătilor situate pe drumul (drumurile) critic(e). Se evaluează o dispersie a valorilor pe care durata proiectului le poate lua prin însumarea dispersiilor duratelor aleatoare ale activitătilor critice. Cu toate că legile de repartitie ale duratelor necesare activitătilor sunt uzual de tipul Beta, teorema limită centrală a calculului probabilitătilor permite asimilarea legii de repartitie a duratei proiectului cu o lege normală cu media si dispersia calculate conform recomandării de mai sus, prin însumare a valorilor analoge ale activitătilor critice.

Această lege normală permite calculul unor probabilităti asociate cu anumite valori ale duratei proiectului recomandate sau chiar impuse.

Structura drumului critic este si ea aleatoare. În raport cu realizarea efectivă a duratelor mai scurte sau mai lungi ale activitătilor, drumul critic poate contine alte si alte submultimi ale multimii de activităti. Există, asadar totdeauna o diferentă între calcul bazat pe durate medii si realitatea executării lucrărilor proiectului.

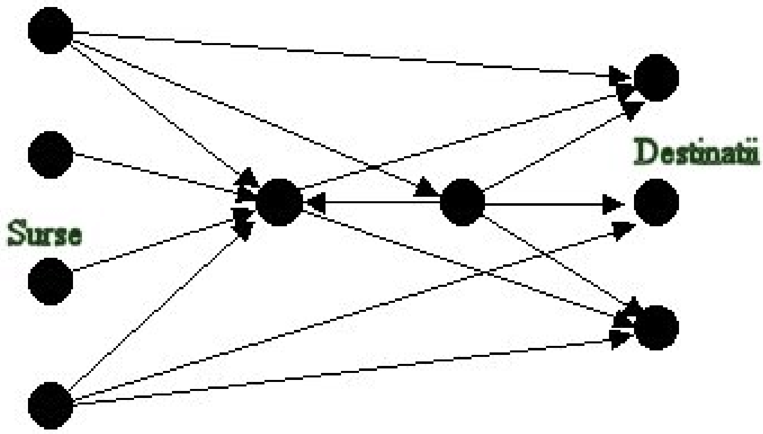
Dacă astfel stau lucrurile, este interesantă pentru cel care conduce lucrările proiectului o sortare a activitătilor în unele care sunt de regulă critice, altele care sunt numai ocazional critice si altele care au sanse mici sau nule de a fi critice. Această utilă sortare se poate face pe baza unor simulări.

Simularea se execută prin generarea aleatoare repetată a unor durate posibile ale activitătilor conform legilor lor de repartitie, de pildă conform unei legi Beta. Fiecare din aceste atribuiri pentru duratele activitătilor a unor valori posibile dar diferit probabile reprezintă o "realizare" posibilă. Pentru fiecare din aceste realizări ipotetice se stabileste drumul critic si se retine de fiecare dată structura drumului critic si lungimea/durata proiectului. În rezultatele acestor simulări (de pildă 100 de simulări) activitătile se vor regăsi pe drumul critic, unele mai frecvent, altele mai putin frecvent (altele, poate, deloc), în general cu frecvente diferite. Se poate spune că unele activităti sunt "mai" critice decât altele si acest "mai" care nuantează aprecierea se pune în relatie cu frecventele de situare a lor pe drumul critic. Dacă o acitvitate apare pe drumul critic în 91% din simulări, ea este dintre cele mai critice. Dacă o acitivitate este critică în 47% din cazuri, ea este mediu critică. Dacă procentul este 3%, este vorba de o activitate numai cu totul ocazional critică. Prin procente de genul exemplificat activitătile se pot ordona în raport cu criticalitatea lor: procentele dau acest grad de criticalitate. Se crează astfel o posibilitate suplimentară de a gestiona mai bine executarea proiectului: o concentrare a resurselor pe activitătile din partea de sus a topului criticalitătii si, gradual o tratare mai putin mobilizantă pentru activitătile din partea de jos a topului.

Duratele de executie a proiectului, diferite de la simulare la simulare sunt un sondaj prin calcul al unei variabile aleatoare despre care s-a spus că în conformitate cu teorema limită centrală a calculului probabilitătilor ar fi o variabilă aleatoare distribuită gaussian (normal), supozitie nu îndeajuns sustinută de realitate.

Fluxuri prin retele

Pentru a ilustra problema generală a fluxurilor prin retele se consideră diagrama de mai jos în care se observă câteva *surse* si un număr de *destinatii*. Tipic, fiecare sursă are o limită superioară de capacitate, fiecare destinatie are o limită superioară a cantitătii de "material" furnizat de surse pe care o poate absorbi.



Între surse si destinatii sunt *noduri* intermediare, în care materialul poate fi încărcat pentru (sau prin care materialul poate curge spre) alte noduri intermediare sau noduri de destinatie, care pot fi, de pildă, niste consumatori. Există în graful prezentat arce, în general orientate, cum sunt figurate aici, dar pot fi si fără orientare si atunci fluxurile pot circula în ambele sensuri. Fiecărui arc îi sunt asociate:

- O limită superioară (numită *capacitate*) a cantitătii de material care se poate scurge într-un mod sau altul pe acel arc
- Un *cost* asociat unitătii de material expediat pe arc

Se pune problema alimentării consumatorilor de la surse, cu un cost minim. Problema este cunoscută ca *problema costului minim asociat fluxului prin retea*.

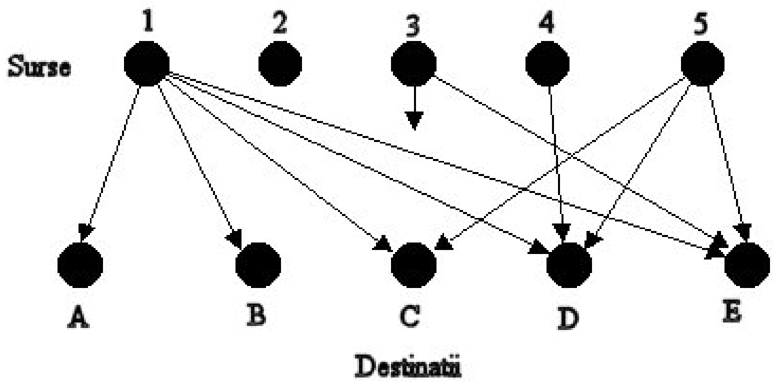
În anii '60 timpurii ai secolului trecut, Ford si Fulkerson au dezvoltat un algoritm pentru această problemă. Algoritmul original a fost de atunci revizuit si îmbunătătit de multe ori, a fost pus pe calculator în multiple editii si variante. Problema în sine este una de programare liniară cu o structură aparte. Astfel de algoritmi specilizati pot rezolva probleme variate. Orice problemă care poate fi reprezentată în forma unui graf ca acela de mai sus poate fi privită ca o problemă de cost minim al fluxului prin retea. Mai

departe sunt prezentate câteva probleme practice potrivite modelării printr-o retea ca aceea de mai sus.
Problema alocării. Fie tabelul de mai jos care contine costurile alocării a 5 sarcini de productie pe 5 executanti.

		Executanti				
		A	B	C	D	E
Sarcini	1	26	16	22	25	30
	2	21	29	33	23	25
	3	28	20	27	32	29
	4	30	19	24	26	24
	5	32	37	30	31	33

Graful asociat acestei probleme este dat în figura alăturată. Pentru claritate, nu sunt cuprinse în graf toate arcele care unesc fiecare sursă (sarcină) cu fiecare destinatie (executant).
Se pune întrebarea: care sarcini trebuie atribuite căror masini pentru a minimiza costul total? Este destul de clar că această problemă poate fi privită ca una de tipul costului minim al fluxului printr-o retea cu particularitătile:

- Fiecare sursă (sarcină) poate furniza o unitate
- Fiecare consumator (executant) poate consuma/executa o unitate
- Fiecare arc are o capacitate de o unitate de flux si un cost conform tabelului de mai sus



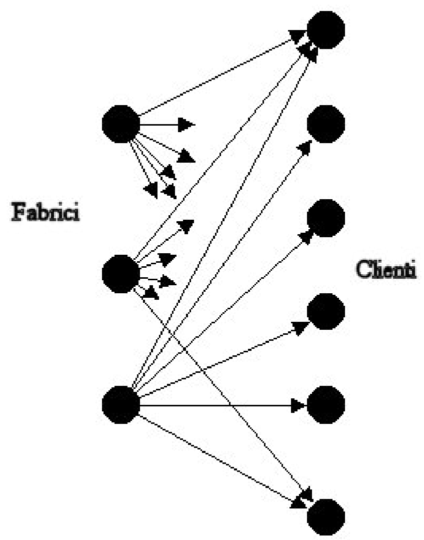
Problemele de acest gen se numesc probleme de *atribuire* (assignare) deoarece ele implică atribuirea a n (aici n = 5) entităti distincte altor m (aici m = 5) entităti distincte. De pildă, în domeniul planificării productiei intereseaza uneori atribuirea de operatori unor masini sau atribuirea de operatori unor operatii sau, analog cu cazul de mai sus, atribuirea unor operatii unor masini.
Problema formulată mai devreme este rezolvată cu calculatorul si solutia este dată mai jos:

Sarcina	Se atribuie executantului:	Costul
1	C	22
2	A	21
3	B	20
4	E	24
5	D	31
Costul total (minim):		118

O problemă de transport. Trei fabrici pot furniza clientilor, în număr de sase, un anumit produs. Cererea de la consumatorii 1, 2, 3, 4, 5 si 6 este de 40, 35, 25, 20, 60 respectiv 30 de tone. Productiile maxime ale celor trei fabrici A, B si C sunt 60, 70 respectiv 80 de tone. Costul de productie pe tonă la cele trei unităti este variabil: 11,3, 11,0 respectiv 10,8 u.m. Costul transportului este dat în tablelul care urmează tot î unităti monetare (u.m.).

		Costul transportului pe tonă la clientul:					
		1	2	3	4	5	6
Fabrica:	A	3,0	2,5	1,8	4,2	3,1	1,5
	B	4,0	1,8	4,6	2,4	3,5	2,2
	C	2,0	2,8	4,8	4,4	1,6	3,6

Se cere să se determine cantitatea de produs livrată de la fiecare fabrică la fiecare client astfel ca costul total să fie minim. Graful asociat problemei este cel alăturat.



Pentru a trata această problemă ca una de minimizare a costului de trecere prin retea este necesar a afla costul pentru fiecare pereche fabrică-client, cost pentru producerea ***si*** transportul unei tone de material de la producător la consumator. Aceste costuri sunt obtinute prin însumarea costului de productie variabil de la fabrică la fabrică cu costurile de transport. Rezultatele sunt tabelate mai jos:

		Costul de productie si de transport pe tonă la clientul:					
		1	2	3	4	5	6
Fabrica:	A	14,3	13,8	13,1	15,5	14,4	12,8
	B	15,0	12,8	15,6	13,4	14,5	13,2
	C	12,8	13,6	15,6	15,2	12,4	14,4

Acum este limpede că problema aceasta poate fi privită si tratată ca una de cost minim al trecerii prin retea:

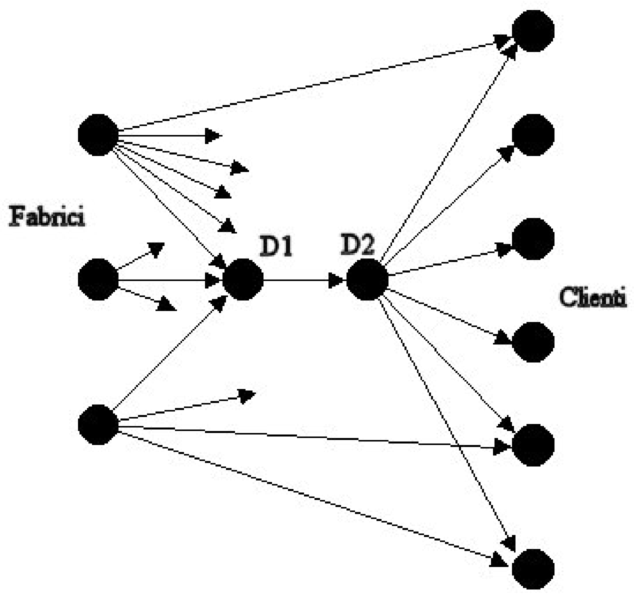
- Fiecare sursă (fabrică) poate debita atât cât îi este productia
- Fiecare consumator (client) are o cerere fixată
- Fiecare arc are o capacitate egală cu cererea clientului la care se adaugă si un cost conform tabelului de mai sus, care contine costul combinat de productie si de transport

Cu datele din tabelul din urmă completate cu cererile clientilor si ofertele producătorilor (însumate, acestea trebuie să fie egale) se obtine rezultatul:

		Se transportă (în tone) la clientul:					
		1	2	3	4	5	6
De la fabrica:	A	20	-	25	-	-	15
	B	-	35	-	20	-	15
	C	20	-	-	-	60	-
Costul total (minim): 2719,50 u.m.							

O problemă de transport cu tranzit. Frecvent, bunurile nu sunt transportate direct de la producător la consumator ci sunt manipulate în puncte intermediare (depozite, magazii). De pildă, la problema precedentă se poate adăuga informatia că s-a creat un depozit nou, intermediar, unde:

- Costul tranzitului este de 0,7 u.m.



- Costul transportului de la fabricile A, B si C la deposit este de 0,1, 0,3 respectiv 0,7 u.m./tonă
- Costul expedierii de la depozit la clientii 1, 2, 3, 4, 5 si 6 este 0,7, 0,9, 1,1, 0,8, 0,6, respectiv 0,9 u.m./tonă

Depozitul poate fi incorporat în retea ca în figura alăturată. S-au adăugat grafului următoarele:

- Două noduri noi, D_1 si D_2; D_1 reprezintă intrarea în depozit ("usa din fată"), D_2 reprezintă iesirea din depozit ("usa din spate")
- Un arc între D_1 si D_2 cu capacitatea egală cu capacitatea totală de productie a fabricilor (maximul de flux ce poate veni de la fabrici la depozit) si cu un cost asociat de 0,7 u.m., care este costul trecerii prin depozit
- Arce de la sursele (A, B, C) la D1, de capacităti egale cu capacitatea de productie a sursei/fabricii si cu costul egal cu suma costurilor de productie si de transport de la fabrică la depozit, adică:
    - Arcul (A, D_1) cu capacitatea 60 si cu costul 11,3 + 0,1 = 11,4
    - Arcul (B, D_1) cu capacitatea 70 si cu costul 11,0 + 0,3 = 11,3
    - Arcul (C, D_1) cu capacitatea 80 si cu costul 10,8 + 0,7 = 11,5
- Arce de la D2 la clientii (1, 2, 3, 4, 5, 6), de capacităti egale cu cererile clientilor si cu costul trimiterii de la depozit la fiecare client, adică:
    - Arcul (D_2, 1) cu capacitatea de 40 si cu costul 0,7
    - Arcul (D_2, 2) cu capacitatea de 35 si cu costul 0,9
    - Arcul (D_2, 3) cu capacitatea de 25 si cu costul 1,1
    - Arcul (D_2, 4) cu capacitatea de 20 si cu costul 0,8
    - Arcul (D_2, 5) cu capacitatea de 60 si cu costul 0,6
    - Arcul (D_2, 6) cu capacitatea de 30 si cu costul 0,9

Problemele de acest tip sunt probleme de transport cu depozitare intermediară. Fluxul de bunuri de la surse la destinatii comportă o transbordare într-o locatie intermediară si nu o livrare directă de la producător la consumator.

Problema se tratează usor prin programare liniară. Dacă variabilele de decizie sunt

- x_i (i = 1, 2, …, 18) cantitătile transferate nemijlocit de la fabrici la clienti pe traseele (A, 1), …, (A, 6), (B, 1), …, (B, 6), (C, 1), …, (C, 6)
- x_i (i = 19, 20, 21) cantitătile transferate mai întâi de la fabrici la depozit (A, D1), …, (A, D1)
- x_i (i = 22) cantitatea manipulată prin depozit, pe arcul (D1, D2)
- x_i (i = 23, …, 28) cantitătile transferate de la depozit la clienti (D2, 1), …, (D2, 6)

atunci, functia obiectiv este functia cost si trebuie minimizată:

$$\begin{aligned}Z = \ & 12.8x_1 + 13.x_2 + 14.4x_3 + 15.5x_4 + 13.8x_5 + 14.3x_6 + 13.2x_7 \\ & + 15.6x_8 + 14.5x_9 + 13.4x_{10} + 12.8x_{11} + 15.0x_{12} + 14.4x_{13} \\ & + 15.6x_{14} + 12.4x_{15} + 15.2x_{16} + 13.6x_{17} + 12.8x_{18} + 11.4x_{19} \\ & + 11.3x_{20} + 11.5x_{21} + 0.7x_{22} + 0.7x_{23} + 0.9x_{24} + 1.x_{25} + 0.8x_{26} \\ & + 0.6x_{27} + 0.9x_{28}\end{aligned}$$

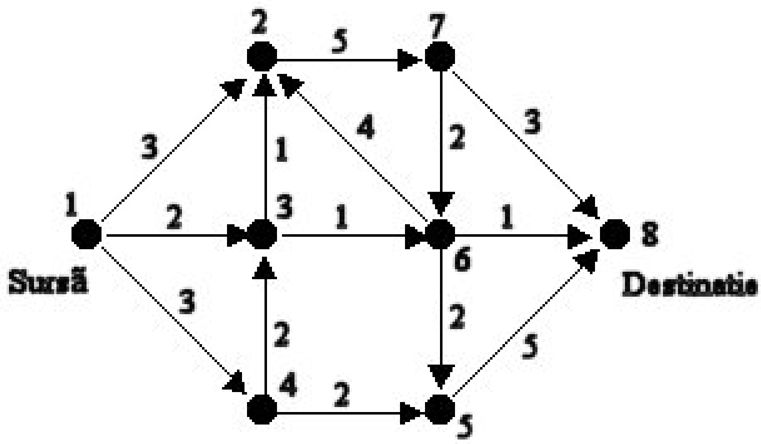
Coeficienti pentru arcele emergente din nodurile sursă (fabricile) reprezintă costurile de productie ale tonei (variabile de la un producător la altul) la care

se adaugă costul transportului pe arcul respectiv. Coeficienti pentru arcele corespunzătoare transferului de la depozit la clienti reprezintă numai costul transportului unei tone. Coeficientul pentru arcul (D1, D2) este costul stocării/manipulării în depozit, de asemenea la tona de produs.
Restrictiile, înafara celor de nenegativtate, sunt în număr de 11.
Primele sunt legate de capacitătile de productie ale celor trei fabrici:
C1 $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_{19} = 60$
C2 $x_7 + x_8 + x_9 + x_{10} + x_{11} + \mathrm{x}_{12} + x_{20} = 70$
C3 $x_{13} + x_{14} + x_{15} + x_{16} + x_{17} + x_{18} + x_{21} = 80$
Următoarele sase sunt legate de capacitătile clientilor de a consuma:
C4 $x_1 + x_7 + x_{13} + x_{23} = 40$
C5 $x_2 + x_8 + x_{14} + x_{24} = 35$
C6 $x_3 + x_9 + x_{15} + x_{25} = 25$
C7 $x_4 + x_{10} + x_{16} + x_{26} = 20$
C8 $x_5 + x_{11} + x_{17} + x_{27} = 60$
C9 $x_6 + x_{12} + x_{18} + x_{28} = 30$
La acestea se adugă două ecuatii de continuitate care exprimă un fapt foarte natural: ce intră în depozit este exact cât se stochează (temporar) si ceea ce se stochează este exact ceea ceea ce iese din depozit cu destinatia clienti.
C10 $x_{19} + x_{20} + x_{21} - x_{22} = 0$
C11 $- x_{22} + x_{23} + x_{24} + x_{25} + x_{26} + x_{27} + x_{28} = 0$
Iată aici un caz în care toate restrictiile sunt egalităti.
Solutia cea mai bună este compusă din valorile $x_{15} = 25$, $x_{18} = 30$, $x_{19} = 60$, $x_{20} = 70$, $x_{21} = 25$, $x_{22} = 155$, $x_{23} = 40$, $x_{24} = 35$, $x_{26} = 20$, $x_{27} = 60$ si toate celelalte variabile de decizie la valori nule.
Pentru această solutie functia obiectiv este minimă în conditiile date în formularea problemei $Z = 2676{,}50$ u.m.
Într-un tabel sintetic, transferurile au loc astfel:

		Către:							
		D1	D2	1	2	3	4	5	6
De la:	A	60	-	-	-	-	-	-	-
	B	70	-	-	-	-	-	-	-
	C	25	-	-	-	25	-	-	30
	D1	-	155	-	-	-	-	-	-
	D2	-	-	40	35	-	20	60	-

Se observă că prin deposit trec 155 unităti (de la D1 la D2). Dacă această capacitate de depozitare/manipulare ar fi limitată, cum se întâmplă adesea, ar trebui introdusă si această conditie restrictivă alături de cele deja formulate.
Se cuvine, poate, a face aici un comentariu. Problemele de transport si de tranzitare pot fi rezolvate relativ usor cu calculatorul. Uneori, se pot formula întrebări suplimentare de genul "dar dacă …" la care se poate răspunde relative rapid, atât la nivelul strategic cât si la cel tactic.
Problema fluxului maxim. O variatie a problemei generale a costului minim al fluxului prin retea este problema stabilirii fluxului maxim care poate fi trimis între o sursă unică (nodul 1) si un consumator unic (nodul 8), ca în

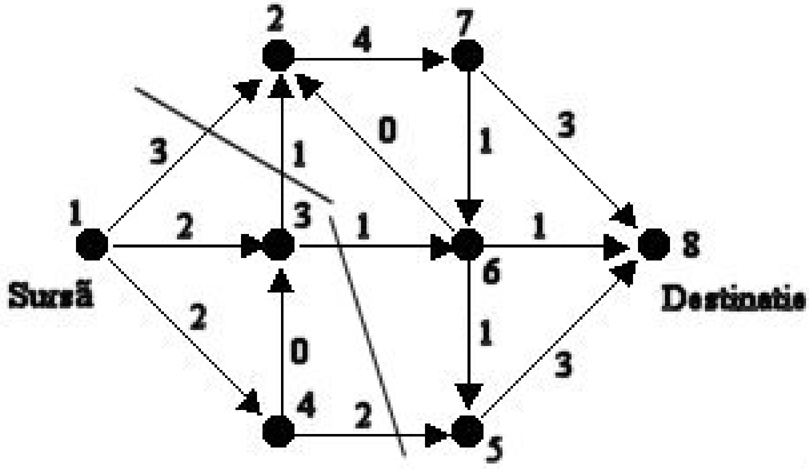
diagrama care urmează, în care fiecare arc are o capacitate (înscrisă în graful lângă arc) care limitează cantitatea scursă prin arc. De data aceasta nu există un cost asociat cu utilizarea arcului, partială sau la capacitate. Arcele sunt orientate în cazul în discutie, dar în general ele pot fi si fără orientare, ceea ce indică posibilitatea de curgere a fluxului în acel arc în ambele sensuri.



O retea de genul acesta poate modela realitatea constituită de un sistem de străzi, sosele sau căi ferate, de linii maritime sau aeriene, un sistem de conducte, poate fi considerat un graf în care nodurile sunt localităti, noduri de comunicatii terestre, maritim-fluviale sau aeriene, puncte de aprovizionare sau de aprovizionat, statii de pompare, iar arcele pot fi tocmai străzile, soselele etc. Capacitătile arcelor pot fi distante, durate necesare parcurgerii lor, costuri specifice de transport etc.
Solutia problemei este reprodusă în tabelul dat mai jos. Se observă un nod fără nici o sosire (1), acesta este nodul sursă, se observă un nod fără nici o plecare (8), acesta este nodul de destinatie, nodul consumator.

		Noduri de sosire							
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Noduri de plecare	*1*		3	2	2				
	2							4	
	3		1				1		
	4					2			
	5								3
	6					1			1
	7						1		3
	8								

Valoarea maximă a fluxului prin retea este de 7 unităti.
Această valoare are proprietatea că se regăseste în bilantul (intrări – iesiri al) oricărei sectiuni care determină de o partitie binară a grafului, partitie în care sursa si destinatia sunt separate. Pe figura de mai jos, în care sunt marcate fluxurile prin valorile-solutie, se poate verifica acest fapt.



Iată si o *teoremă*, a *tăieturii minime si a fluxului maxim*, care spune că fluxul maxim posibil este egal cu capacitatea tăieturii minime care deconectează sursa de consumator. În figura de mai sus tăietura minimă corespunde sectionării arcelor (1, 2), (3, 2), (3, 6) si (4, 5) – eliminarea acestor arce face ca sursa si destinatia să fie separate si capacitatea lor totală este 7.

Problema *drumului minim* într-un graf este de foarte mare actualitate în tările cu un standard de viată ridicat. Este o problemă de ordin public. După cum si numele problemei indică, este interesant pentru cel care călătoreste sau transportă bunuri sau persoane a alege din mai multe drumuri posibile între două localităti sau între două puncte ale unei aceleiasi localităti drumul cel mai scurt. Această optimizare este încurajată prin afisarea pe Internet a unor hărti care pot fi dilatate (zoom in) sau contractate (zoom out) la o mărime a detaliilor adecvată. În spatele acestor grafuri-hărti sunt algoritmi de stabilire a drumului cel mai scurt între punctul de plecare si punctul de destinatie. Algoritmul *Ford* este unul dintre acestia. Este o problemă clasică (re)adusă în actualitate de mobilitatea foarte ridicată a omului contemporan.

ELEMENTE DE TEORIA DECIZIILOR

Algoritmi de decizie simpli

Se presupune că într-un context dat sunt m decizii posibile d_1, d_2, …, d_m. Sistemul economic poate fi în n stări distincte s_1, s_2, …, s_n. Dacă se ia decizia d_i si sistemul este în starea s_j decizia este acompaniată de profitul sau de penalitatea $r(i, j)$. Numerele $r(i, j)$ se pot aseza într-o matrice cu m linii si n coloane, numită si matricea de plăti. Se pune problema luării deciziei optime. Un exemplu e continut în tabelul care urmează:

Decizii\Stări	s_1	s_2	s_3	s_4	s_5
d_1	3	1	3	2	6
d_2	6	7	–5	8	0
d_3	3	4	–1	–2	9
d_4	3	3	2	–2	–1

Există mai multe metode de a decide.
Metoda maxmin are în vedere câstigul sau profitul sigur. Pe baza tabelului de plăti dat mai sus se calculează profitul minim pentru fiecare decizie în parte. Rezultă ceea ce este scris în tabelul prezentat imediat.

Decizii	Profitul minim
d_1	1
d_2	–5
d_3	–2
d_4	–2

Decizia vizează maximul profitului minim adică decizia d_1.
Metoda Laplace se bazează pe principiul asa-numitei motivări insuficiente. Oricare dintre stările sistemului este posibilă cu probabilităti egale si se urmăreste profitul maxim sau pierderea minimă.
O mediere pe orizontală produce tabelul următor.

Decizii	Profitul mediu
d_1	3,0
d_2	3,2
d_3	2,6
d_4	1,0

si decizia cea mai potrivită (d_2) se vede imediat.

Metoda *regretului minimax* are în vedere diferenta dintre cel mai mare profit pentru o stare dată si profitul real pentru fiecare decizie luată. Decizia este luată pe principiul minimizării regretului maxim posibil. Pentru a decide se construieste următoarea matrice de regret, care corespunde matricii de plăti de mai sus:

Decizii\Stări	s_1	s_2	s_3	s_4	s_5
d_1	3	<u>6</u>	0	<u>6</u>	3
d_2	0	0	8	0	9
d_3	3	3	4	10	0
d_4	3	4	1	10	10

Decizia cea mai bună este d_1 deoarece regretul maxim este cel mai mic fata cel de la celelalte decizii.
Metoda *Hurwicz* introduce si o notă subiectivă în procesul decizional, printr-un asa-zis *grad de oprimism*. Gradul de optimism este un coeficient subunitar si pozitiv, α. În cazul unei decizii d_i optimismul se raportează la profitul maxim $r(i, j)$, iar pesimismul la profitul minim $r(i, j)$. Criteriul din metoda Hurwicz combină liniar cu coeficientii α, respectiv $(1 - \alpha)$ cele două valori extreme ale profitului. Decizia se ia pe valoarea maximă a acestei combinatii liniare convexe. De pildă, pe matricea de plăti dată mai devreme si cu un coeficient de optimism $\alpha = 0{,}8$, se poate elabora tabelul de mai jos, care serveste la a decide.

Decizii	Profit maxim $r_{\max}$	Profit minim $r_{\min}$	$\alpha r_{\max}+(1-\alpha)r_{\min}$
d_1	6	1	5,0
d_2	8	–5	5,4
d_3	9	–2	6,8
d_4	3	–2	2,0

Decizia optimă este si de data aceasta imediat vizibilă.
O cuprindere într-un tabel unic a deciziilor si profiturilor obtenabile în fiecare caz, pentru fiecare din cele patru metode

Metoda	Decizia optimă	Profit
Maxmin	d_1	1
Laplace	d_2	3,2
Regret minmax	d_1	6
Hurwicz	d_3	6,8

aduce o umbră de îndoială asupra procesului decizional: metode diferite duc la decizii diferite, cu valori ale profitului, de asemenea, foarte diferite. Aplicarea metodelor mentionate are însă totdeauna si o nuantă strategică, de context economic. Acest element apare explicit în cazul metodei Hurwicz.

Arbori decizionali

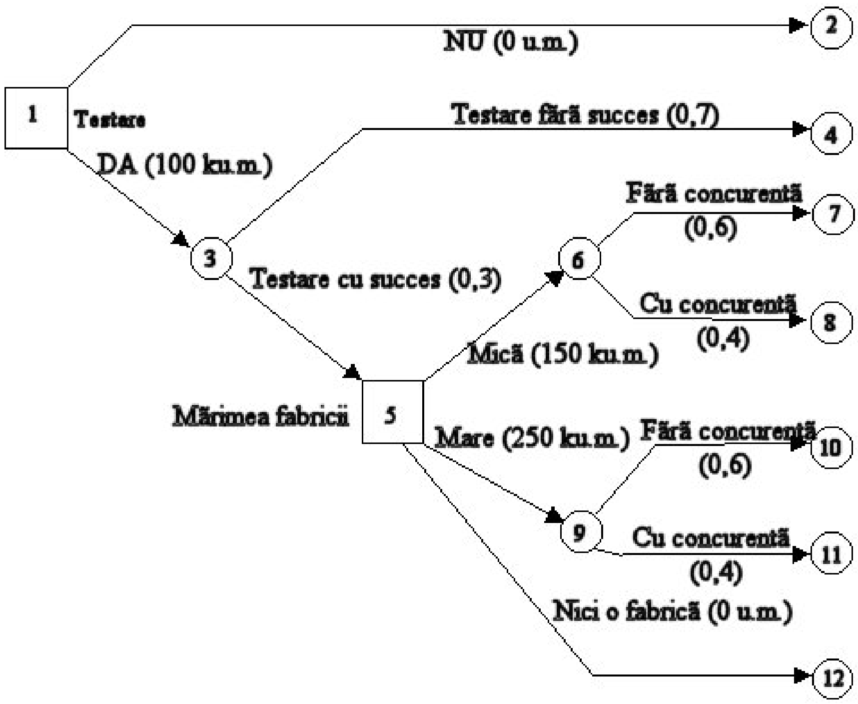
În multe probleme din practica conducerii sistemelor de productie, sansa – un alt nume pentru probabilitate – joacă un rol important. Analiza decizională este termenul general asociat metodelor de a analiza probleme care comportă riscuri/incertitudini/probabilităti. Arborii decizionali constituie una din metodele specifice analizei decizionale si un exemplu o va ilustra.

Exemplu. O companie are de hotărît asupra unui produs dezvoltat într-unul din laboratoarele de cercetare proprii. Trebuie să decidă dacă va proceda sau nu la testarea pietii pentru acel produs, de aici încolo notat cu **P**. Se estimează că marketingul etapei de testare va costa 100.000 unităti monetare (u.m.). În ceea ce priveste succesul produselor în testarea pietii, experienta anterioară indică o sansă de (numai) 30%.

Dacă **P** are succes în etapa de testare a pietii atunci compania mai are de luat o hotărîre relativ la dimensiunea fabricii care urmează a produce noul produs. Construirea unei unităti mici ar costa 150.000 u.m. si ar produce 2000 unităti fizice (u.f.) pe an. O fabrică mare ar costa ca investitie 250.000 u.m. si ar produce 4.000 u.f. pe an.

Departamentul de marketing a estimat că sunt sanse de 40% ca concurenta să răspundă cu un produs similar si că pretul pe fiecare unitate vândută (în u.m.) ar fi, dacă productia se vinde integral, după cum urmează:

	Unitate mică	Unitate mare
Cu răspuns al concurentei	20 u.m.	35 u.m.
Fără răspuns de la concurentă	50 u.m.	65 u.m.



Se admite că cererea pentru **P** va exista pe o perioadă estimată la 7 ani si că anual, costul functionării unitătii productive este de 50.000 u.m., pentru a usura calculele, acelasi, indiferent de dimensiunea unitătii de productie. Întrebarea este dacă compania trebuie sau nu să meargă înaite cu testarea pietii pentru **P**. Obtinerea răspunsului parcurge procedura prezentată în continuare.
Cu toate că exemplul propus este în bună măsură simplificat, el reprezintă deplin tipul de decizii de luat adesea, când este vorba de produsele noi. În particular, este de observat că decizia de a testa piata nu poate fi separată de alte decizii viitoare asupra eventualei profitabilităti dacă testul de piată pentru **P** este un succes.
Pentru a abilita decidentul (în particular, studentul) să vadă ce se întâmplă, se consideră figura de mai sus în care este reprezentat arborele decizional al problemei. În graful de mai sus se disting două tipuri de noduri: noduri decizionale (reprezentate prin dreptunghiuri) si noduri probabilistice (reprezentate prin cercuri). Mai sunt nodurile terminale care se regăsesc în partea cea mai din dreapta a grafului cu structură arborescentă.
Nodurile de decizie reprezintă în parcurgerea de la stânga la dreapta puncte în care compania trebuie să aleagă una din mai multe posibilităti. La primul nod decizional, de pildă, compania trebuie să decidă "a renunta la produs" sau "a-l testa pe piată".
Nodurile probabilistice reprezintă puncte în care probabilitătile sau sansele (marcate pe arcele energente) joacă un rol dominant si înmănunchiază situatii, evenimente asupra cărora compania nu are efectiv nici un control.
Nodurile terminale reprezintă finalul drumurilor posibile de la stânga la dreapta prin arborele decizional.
Partea dificilă a acestei metode a arborelui decizional este aceea de a trasa o diagramă de genul celei de mai sus pornind de la varianta scrisă, descriptivă a problemei. Odată parcursă această etapă, procedura este destul de simplă si directă. Trebuie spus că un arbore decizional nu începe totdeauna cu un nod decizional. La trasarea unui arbore decizional, cel care elaborează arborele trebuie să se întrebe repetat "Ce se poate întâmpla în continuare?", la fiecare nod al arborelui, pe măsură ce el este asezat pe grafic.
Se observă în graf includerea posibilitătii "nici o fabrică" în nodul decizional relativ la dimensiunile fabricii. Aceasta incluziune este *necesară* deoarece este posibil ca investitia într-o fabrică mică sau mare să nu fie profitabilă chiar dacă testul de piată a fost un test reusit. Este o practică obisnuită în problemele tratate cu ajutorul arborilor decizionali să se includă în graf la nodurile decizionale, posibilitatea unei decizii de genul "a nu întreprinde nimic" care este totdeauna o alegere implicită.
O atentie trebuie acordată structurii grafului: arborele de decizie trebuie desenat astfel ca de la nodul initial la oricare dintre nodurile terminale să existe un *drum* unic.
Pentru a facilita discutia pe arborii deicizionali, nodurile, fie ele decizionale, probabilistice sau terminale se numerotează. În exemplul curent, s-au numerotat cu 1, 2, …, 12. Pentru fiecare nod probabilistic se recomandă o numerotare a posibilitătilor: la nodul 1 sunt posibile două situatii, care ar

putea fi marcate cu 1 si 2, iar la nodul 5 sunt de luat în considerare trei posibilităti, care ar putea primi numerele 3, 4 si 5.

Arborele decizional ajută, fără îndoială la formarea unei viziuni mai clare asupra naturii problemei. Deocamdată, însă, nu a dat un răspuns la întrebarea primă, generatoare a problemei: a renunta sau a proceda la testarea pietei pentru produsul **P.** Pentru a obtine răspuns la această întrebare sunt de parcurs în continuare încă doi pasi descrisi mai departe. În acesti pasi sunt necesare informatii numerice relativ la vânzările viitoare, la preturi, la costuri etc., cu toate că nu totdeauna sunt accesibile numere exacte relativ la acestea. Decizia de a testa sau nu produsul pe piată s-ar putea schimba la modificarea acestor valori. O analiză a sensibilitătii deciziei poate si chiar trebuie să fie făcută de îndată ce s-au efectuat calculele de bază utilizând valori mai mult sau mai putin ipotetice.

Pasul 1 este pasul în care, pentru fiecare cale prin arbore, de la rădăcină (nodul intial) la un nod terminal (frunză) al unei ramuri, se evaluează profitul asociat acelei căi. În esentă în acest pas se parcurge diagrama de la stâga la dreapta.

Calea la nodul terminal 2 – se renuntă la **P**.

Recuperări = 0

Costuri = 0

Profit = 0

Aici, ca si în continuare, se ignoră orice sumă cheltuită deja în faza de dezvoltare a produsului (acela este un cost care nu mai poate fi modificat indiferent ce decizii vor fi luate în viitor si, logic, nu joacă nici un rol în procesul decizional.

Calea la nodul terminal 4 – se testează piata pentru produsul **P**, se constată că nu este un succes si se renuntă la el.

Recuperări = 0

Costuri = 100.000

Profit = – 100.000

Calea la nodul terminal 7 – se testează piata (cost 100.000 u.m.), **P** este un succes, se construieste o unitate de productie mică (cost 150.000 u.m.) si concurenta lipseste (venitul pe 7 ani la o productie de 2.000 unităti fizice (u.f.) pe an, la un pret de 65 u.m. per bucată = 910.000 u.m.)

Recuperări = 910.000

Costuri = 250.000 + 7x50.000 (costul de operare)

Profit = 310.000

Calea la nodul terminal 8 – se testează piata pentru **P** (cost 100.000 u.m.), produsul este un succes, se construieste o fabrică mică (cost 150.000 u.m.) si concurenta este prezentă (venitul pa 7 ani la 2.000 de u.f. anual, vândute cu 35 u.m. per bucată = 490.000 u.m.)

Recuperări = 490.000

Costuri = 250.000 + 7x50.000

Profit = – 110.000

Calea la nodul terminal 10 – se testează piata (cost 100.000 u.m.), produsul este de succes, se construieste o fabrică mare (cost 250.000 u.m.) si concurenta lipseste (venitul pe 7 ani la 4.000 de u.f. anual si la pretul de vânzare de 50 u.m. = 1.400.000 u.m.)

Recuperări = 1.400.000
Costuri = 350.000 + 7x50.000
Profit = 700.000
Calea la nodul terminal 11 – se testează piata (cost 100.000 u.m.), produsul este de succes, se construieste o fabrică mare (cost 250.000 u.m.) si concurenta este prezentă (venitul pe 7 ani la 4.000 de u.f. anual si la pretul cu amănuntul de 20 u.m. = 560.000 u.m.)
Recuperări = 560.000
Costuri = 350.000 + 7x50.000
Profit = – 140.000
Calea la nodul terminal 12 – se testează piata (cost 100.000 u.m.), produsul este de succes, se decide a nu construi nici o capacitate de productie.
Recuperări = 0
Costuri = 100
Profit = –100
Această din urmă cale include decizia "a nu întreprinde nimic" în ceea ce priveste investitia într-o capacitate de productie. Este o decizie justificată în multe împrejurări.
Cu rezultatele de mai sus se poate constitui tabelul următor care arată pentru fiecare ramură profitul asociat cu acea ramură, de la nodul initial (rădăcină) până la fiecare nod terminal (frunză).

Nodul terminal	Profit
2	0
4	–100
7	310
8	–110
10	700
11	–140
12	–100

Deocamdată probabilitătile nu au fost utilizate în rezolvarea problemei. Se va face aceasta în pasul al doilea în care se lucrează pe arbore de la dreapta la stânga.
Pasul 2. Se ia nodul 6, un nod cu probabilitate (evenimential) care are ramuri către nodurile terminale 7 si 8. Ramura la nodul terminal 7 se produce cu o probabilitate de 0,6 si aduce un profit total de 310.000 u.m., iar ramura la nodul terminal 8 se produce cu probabilitatea de 0,4 si profitul total este –110.000 u.m. În consecintă, *valoarea asteptată în bani* (EVM – Expected MonetaryValue), în sensul m ediei variabilei aleatoare *profitul după nodul 6* este
0,6 x (310.000) + 0,4 x (–110.000) = 142.000 u.m.
În esentă această valoare reprezintă, asadar, profitul asteptat (mediu) din acest nod probabilistic. Această EVM asociată unui nod probabilistic este "suma pe toate ramurile cu originea în acel nod, a produselor dintre probabilitătile si valorile monetare asociate ramurilor respective". Continuând, EMV pentru nodul 9 cu ramuri către nodurile terminale 10 si 11 este

0,6 x (700.000) + 0,4 x (–140.000) = 364.000 u.m.

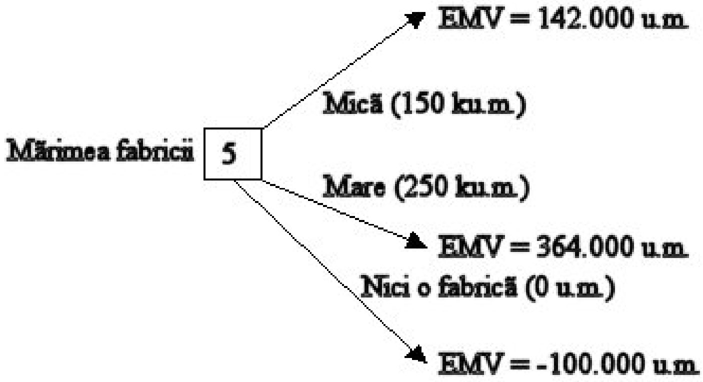
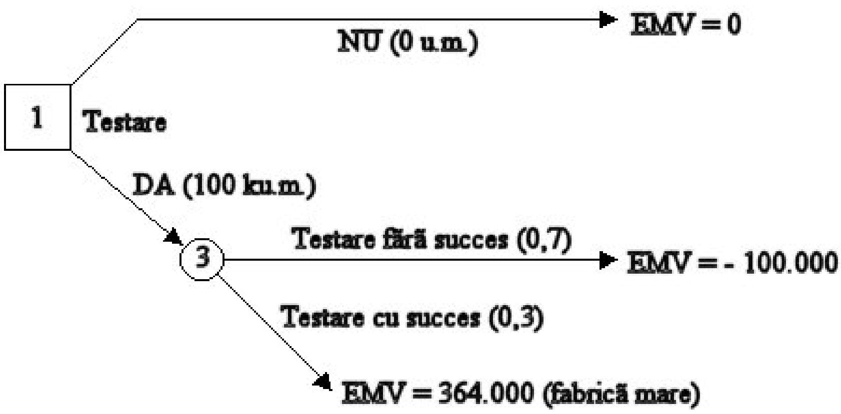


Figura de mai sus reia nodul decizional unde se hotărăste dimensiunea fabricii. Se poate vedea cum nodurile probabilistice au fost substituite cu EMV-urile lor.
După cum se vede, nodul decizional relativ la dimensiunea fabrici este ramificat pe trei posibilităti: fabrică mică, EMV = 142ku.m., fabrică mare, EMV = 364ku.m. sau nici o fabrică, EMV = – 100ku.m. Este limpede că în termeni monetari calea mijlocie este cea mai atractivă si, de aceea, se poate renunta la clelalte două. Asfel modificată problema poate fi reprezentată printr-o variantă revizuită a arborelui decizional, care este reprezentat în figura care urmează.



Acum se poate repeta procesul efectuat mai devreme. EVM pentru nodul probabilistic 3, care se leagă de succesul sau lipsa de succes a testului de piată este dată de
0,3 x (364.000) + 0,7 x (– 100.000) = 39.200 u.m.
Asadar, în nodul de decizie referitor la a testa sau nu produsul pe piată, sunt două posibilităti: renutare la **P**, EMV = 0 sau testarea lui **P** pe piată, EMV = 39.200 u.m. Este clar că în termeni monetari alternativa a doua este preferabilă celeilalte si se decizia este de a testa piata.
În sumar, trebuie evidentiat clar ce s-a decis în urma procesului descris mai sus: trebuie testată piata pentru **P** si această decizie are o valoarea în bani asteptată, medie (EVM) de 39.200 u.m.; dacă **P** este un succes de piată

atunci se anticipează la această etapă că este potrivită construirea unei fabrici mari (rezultatul deciziei din nodul relativ la dimensiunea capacitătii de productie).
De retinut că EVM pentru decizia cerută de problemă, 39.200 u.m. nu reflectă valoric ceea ce se va întâmpla efectiv dacă se continuă cu constrirea unei fabrici etc. Această valoare este mai curând o medie sau o sperantă matematică, cum se mai spune, ca si când ar fi vorba de producerea mai multe ori a fiecăreia din cele trei posibilităti. De fapt, din cele trei posibilităti poate apărea numai una o singură dată. Dacă se urmează calea de mai sus de a testa piata pentru **P** atunci rezultatul real în bani poate fi unul din sase, [–100, 310, –110, 700, –140, –100] (în mii de u.m.), corespunzătoare nodurilor terminale 4, 7, 8, 10, 11 si 12 în functie de deciziile viitoare si de probabilitatea producerii unor evenimente.
Conceptual, nodurile terminale sunt imaginate ca situatii care pot fi atinse ca urmare a deciziei de a testa piata, ca rezultate ale unui set de scenarii posibile. Ca urmări ale deciziei de a testa piata, cel mai bun rezultat posibil este un plafon superior (*upside*) si cel mai slab rezultat posibil este o limită de jos (*downside*). Trebuie avută aici o oarecare grijă deoarece, cum s-a spus, rezultatul real în bani va depinde atât de deciziile viitoare cât si de sansa de producere a unor evenimente viitoare. Dacă se angajează investitia într-o fabrică mare (presupunând că testul de piată a fost reusit) atunci multimii de scenarii posibile îi corespunde multimea de rezultate {100.000, 700.000, –140.000} asociate nodurilor term inale 4, 10 si 11 si, deci, plafonul superior este de 700.000 u.m. iar limita minimală este de –140.000 u.m.
Înainte de a decide asupra investitiei într-o fabrică mare (tot în cazul testului de piată reusit), multimea de perspective în termeni bănesti este {100.000, 310.000, –110.000, 700.000, –140.000, –100.000} corespunz ătoare nodurilor 4, 7, 8, 10, 11 si 12. Plafonul superior este de 700.000 u.m., iar limita minimală este de –140.000 u.m.
În exemplul în discutie, plafoanele superioare si inferioare sunt la fel, indiferent dacă se optează pentru o fabrică mare sau nu. Diferă numai lista de scenarii prin numărul de scenarii cuprinse.
Deoarece calculul pe arborele decizional este atât de direct, este relativ usor a conduce o analiză a sensibilitătii pentru a vedea cum se schimbă succesiunea actiunilor dacă datele problemei se schimbă.
Analiza sensibilitătii. Se consideră arborele decizional dat mai sus. Este evident că decizia de a testa piata este influentată de profitul de 700.000 u.m. obtinut dacă testul de piată este un succes si dacă se alege varianta de a construi o fabrică mare si concurenta nu există. Dar numărul 700.000 poate varia sau probabilitatea ca acest rezultat să se producă se poate modifica. Schimbă aceste variatii decizia de a testa piata?
Se poate presupune, de pildă, că probabilitatea lipsei de concurentă în conditiile construirii unei fabrici mari nu mai este 0,6 ci 0,45. Aceasta implică o probabilitate de prezentă a concurentei de 0,55 = 1 – 0,45. Prin refacerea calculelor pe arborele decizional se obtin rezultate diferite, dar decizia de a testa piata este încă optimă.

Analiza de sensibilitate se poate conduce într-un mod mai sistematic: se atribuie probabilitătii legate de absenta concurentei un simbol p si se lucrează asupra expresiilor pentru EMV. Asadar, probabilitatea lipsei de replică din partea concurentei nu mai este 0,6 ci p. Complementara, $1 - p$ este probabilitatea de a avea concurentă. Se presupune că, pentru o fabrică mică, probabilitătile concurentei/lipsei de concurentă rămân ca mai înainte. Este vizibil că pe măsură ce p descreste, la o anumită valoare este de preferat o instalatie mică în locul uneia mari (de pildă, la extrem, dacă $p = 0$ atunci o fabrică mică cu EMV de 142.000 u.m. este preferabilă unei fabrici mari pentru care EMV este –140.000 u.m.). Prin urmare, se poate pune întrebarea legitimă: "*Cât de mică trebuie să fie probabilitatea **p** înainte de a prefera o fabrică mică*?"

Răspunsul îl dă situatia în care optiunea între o instalatie mare si una mică este indiferentă deoarece EMV-urile sunt egale, ceea ce se produce atunci când

$$p(700) + (1 - p)(-140) = 142$$

ceea ce-i tot una cu

$$840p - 140 = 142$$

o ecuatie simplă cu solutia $p = 282/840 = 0,3357$.

Asadar, dacă p scade sub 0,3357 capacitatea de productie mică va fi preferată celei mari. Acest tip de analiză sistematică a sensibilitătii poate fi în anumite împrejurări preferabilă încercării simple a unor numere diferite si efectuarea de fiecare dată a calculelor pentru a vedea efectul.

Variatiuni pe tema arborilor de decizie. Tehnica arborelui decizional prezentată mai sus poate fi aplicată si în alte împrejurări, în raport cu alte principii si criterii. Există câteva variante ale acestei tehnici dintre care unele sunt prezentate pe scurt imediat mai jos.

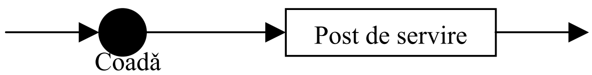
În exemplul discutat mai sus, s-au estimat sume de bani primite pe durata a 7 ani. Este stiut că o sumă primită în 7 ani este de regulă mai putin valoroasă decât aceasi sumă primită azi. O tehnică numită *discounting* sau *discounted cash flow* (care se referă la *valoarea netă prezentă* a oricărei sume de bani) poate fi utilizată pentru a cuprinde si a depăsi acest inconvenient. Aplicând discount-ul se modifică de fapt numerele atasate arborelui decizional astfel ca evaluările să se facă pe baza unui echivalent monetar actualizat. Aceasta nu afectează procesarea arborelui care rămâne exact cel de mai sus.

Tot în exemplul dat mai sus s-a calculat o valoare pentru fiecare nod evenimential. S-a utilizat EMV ca valoare asociată nodului dar această valoare este în unele privinte arbitrară. Specialistii au sugerat alte modalităti de a calcula valoarea atasată unui nod eveniment. În termeni mai clari, nu există vreo lege care să garanteze că valoarea unui nod eveniment trebuie să fie egală cu valoarea EMV. Dimpotrivă, EMV este o valoare medie si într-un nod eveniment nu se va observa niciodată media ci ceva care se întâmplă numai o dată deoarece în nodul eveniment 6 competitia există sau nu există. De aceea media poate fi înselătoare si e necesară o tratare diferită a oricărui nod eveniment. Dacă pierderea de bani ar putea fi inacceptabilă si dacă procesul de decizie este afectat de un conservatorism de înteles, nodului eveniment i s-ar putea atribui rezultatul cel mai slab posibil. O asemenea

strategie este una pesimistă (o astfel de tratare a problemei în nodul 6 ar aduce valoarea –110.000 si nu EMV-ul de 142.000).
O strategie alternativă ar putea fi una optimistă: calculul valorii din nodului eveniment i s-ar asocia rezultatul cel mai bun posibil (în nodul 6 s-ar pune valoarea 310.000 si nu EMV = 142.000).
O altă strategie de luat în considerare ar consta în asocierea nodului eveniment cu valoarea cea mai probabilă. O asemenea strategie ar atribui nodului 6 o valoare de 310.000 u.m. si nu aceea de 142.000 u.m. care este EMV.
Ca strategie intermediară se poate lua ca valoare pentru un nod-eveniment o combinatie ponderată a EMV si a valorilor din strategiile optimistă si pesimistă. Literatura oferă o varietate mare de retete de asociere a unei valorii cu un nod-eveniment.
În nodurile de decizie, mai sus se alege una din posibilităti pe baza unei reguli implicite “se alege cel mai mare EMV”. Dar mai sunt si alte reguli bune, egal utilizabile, de pildă “se alege cel mai bun raport profit/investitie” (ROI – *return of investment*).
Iată ce se obtine dacă se reiau în consideratie capacitătile de productie mică si mare din cazul de mai sus. O fabrică mică conduce la o EMV (de fapt profitul net asteptat) de 142.000. Este implicată o investitie de 100.000 pentru testarea pietii si 150.000 pentru constructie, asa încât rezultă un ROI = 142.000/(100.000 + 150.000) = 0,568.
O fabrică mare conduce la o EMV (iarăsi profitul net asteptat) de 364.000. Investitia de 100.000 pentru testarea pietii si de 250.000 pentru construirea unitătii face un ROI de 364.000/(100.000 + 250.000) = 1,04. Cu toate că noul criteriu, ROI maxim, conduce la aceeasi decizie relativ la dimensiunea capacitătii de productie, în alte situatii, schimbarea criteriului poate conduce la o decizie diferită. De pildă, dacă la nodul 9 ar fi trecută valoarea 175.000 atunci pe baza EMV la nodul decizional 5 s-ar alege tot o fabrică mare. Dar pe baza ROI [175.000/(100.000 + 250.000) = 0,5] s-ar hotărî construirea unei fabrici mici.
Un alt aspect care dă relief utilizării arborilor decizionali este descris în continuare. Prin folosirea într-un arbore decizional a valorilor în bani se obtine, de pildă, că o pierdere de 200.000 u.m. este de două ori mai grea decât o pierdere de 100.000 u.m. Dacă compania nu are de unde să piardă 200.000 dar are de unde acoperi 100.000 u.m. atunci este clar că pierderea de 200.000 este considerată mult mai grea decât una de 100.000 u.m. În supliment, deciziile sunt făcute de oameni din companie; compania face profitul si pierderile, nu oamenii care decid.
Astfel, ideea de “utilitate” (*utility*) constă în a înlocui valorile în bani la fiecare nod terminal în *puncte de utilitate*, care reflectă vederile decidentului (sau companiei) asupra acelor sume de bani (de exemplu, o pierdere de 100.000 u.m. poate echivala cu –5 puncte de utilitate, iar o pierdere de 200.000 poate fi asociată cu –500 de astfel de puncte). În termeni directi, valorile bănesti sunt înlocuite cu puncte. Traducerea valorilor în puncte este un proces imprecis dar înlocuirea abilitează decidentul/decidentii a avea o a doua vedere asupra fondurilor în raport de importantă. Odată stabilite valorile utilitare, se procedează în continuare ca mai sus.

SISTEME CU ASTEPTARE

Un număr considerabil de sisteme de productie functionează în virtutea raportului client-server, adică sunt prestate anumite servicii în folosul unor clienti generici care pot fi persoane, echipamente, subansamble etc. Clientii solicită serviciul (uneori mai multe servicii) după reguli aleatoare. Servirea însăsi necesită de cele mai multe ori o durată de executare aleatoare, care nu poate fi anticipată decât în probabilitate.
În sistemele cu asteptare se nasc uzual cozi sau linii (fire) de asteptare. În dimensionarea unui sistem cu asteptare trebuie să se realizeze pe criterii economice un echilibru între durata asteptării pentru obtinerea serviciului sau serviciilor solicitat(e) (asteptarea costă) si numărul de posturi de servire (operarea fiecărui post de servire costă, de asemenea).
În principiu, un sistem cu asteptare oricât de complicat poate fi descompus în subsisteme în care se rezolvă o singură problemă, un anumit gen de serviciu solicitat de client, din mai multe posibile. Schema acestui subsistem este ilustrată în figura imediat următoare.



Pentru a analiza un asemenea (sub)sistem sunt necesare informatii relativ la sosiri, la mecanismul servirilor si la caracteriticile cozii.

Procesul sosirilor

Procesul sosirilor se descrie prin:
- Modalitatea sosirii clientilor – individual sau în grup(e)
- Repartizarea în timp a sosirilor – o lege de repartitie a timpului între două sosiri consecutive
- Cardinalul multimii de clienti – număr finit sau infinit de clienti

Sosirile cele mai simple si mai comod de tratat sunt cele la intervale regulate, egale.
Un flux de clienti poissonian corespunde unor sosiri aleatoare. În acest caz duratele între două sosiri succesive se distribuie după o lege exponentială. Modelul Poisson este important deoarece se potriveste multor situatii din lumea reală. (Sub)sistemul este caracterizat pe deplin în ceea ce priveste sosirile prin rata medie a sosirilor.

Sosirile pot fi planificate, pot fi grupate, pot fi cu rată variabilă în timp (de pildă rate diferite în momente diferite ale zilei, ale săptămânii, ale anului etc.).
În cazul aleator, probabilitatea sosirii în sistem a unui client, în intervalul scurt Δt este practic proportională cu timpul

$$\lambda \Delta t + O(\Delta t^2)$$

Factorul λ este o frecventă medie (rata medie) a sosirilor/solicitărilor, iar termenul $O(\Delta t^2)$ notează o corectie a valorii probabilitătii, care tinde către zero la fel de repede ca si Δt^2. Evenimentul contrar, *non*-sosirea are probabilitatea complementară

$$1 - \lambda \Delta t - O(\Delta t^2)$$

Sosirea concomitentă a mai multor clienti este considerată cu totul improbabilă. De ce? Pentru că probabilitatea unei sosiri este un număr pozitiv subunitar ca orice probabilitate si este mic dacă intervalul Δt este ales suficient de mic. Sosirile sunt evenimente independente si probabilitatea producerii a două (sau mai multe) sosiri în intervalul Δt este produsul probabilitătilor individuale ale acelor sosiri, adică o putere a unui număr mic. Puterea (întreagă su pozitivă a) unui număr subunitar mic este, desigur, un număr si mai mic. Probabilitatea aceasta foarte mică poate fi asociată practic cu imposibilitatea producerii concomitente a două sau mai multe sosiri într-un acelasi interval Δt.
Se pune acum problema a se calcula probabilitatea ca într-un interval mai îndelungat $h = m\Delta t$ să se producă r ($r \le m$) sosiri în sistem. Cele r sosiri pot fi situate pe axa timpului în moduri variate în succesiunea celor m intervale de durată Δt, adică pot fi intercalate sau nu cu *non*-sosiri. Dacă se iau în calcul toate posibilitătile de sosire (unică)/*non*-sosire în cele m intervale, se obtine

$$p_r = \frac{m!}{r!(m-r)!}(\lambda\ \Delta t)^r (1-\lambda\ \Delta t)^{m-r}$$

ceea ce după înlocuirea $\Delta t = h/m$ devine

$$p_r = \frac{m!}{r!(m-r)!}\frac{(\lambda h)^r}{m^r}\left(1-\lambda\frac{h}{m}\right)^{m-r}$$

Luând limita când $\Delta t \to 0$, fapt echivalent cu trecerea $m \to \infty$, se obtine

$$\lim_{m\to\infty} p_r = \frac{(\lambda h)^r}{r!} \lim_{m\to\infty} \frac{m!}{m^r(m-r)!}\left(1-\lambda\frac{h}{m}\right)^{m-r}$$

Dar

$$\lim_{m\to\infty} \frac{m!}{m^r(m-r)!} = \lim_{m\to\infty}\left(1-\frac{1}{m}\right)\ldots\left(1-\frac{r-1}{m}\right) = 1$$

si

$$\lim_{m\to\infty}\left(1-\lambda\frac{h}{m}\right)^{m-r} = e^{-\lambda h}$$

Asadar, probabilitatea căutată este dată de expresia

$$p_r = \frac{(\lambda\, h)^r}{r!} e^{-\lambda\, h}$$

care nu este altceva decât probabilitatea dată de o lege de repartitie poissoniană. Între două sosiri succesive se scurge un timp t în care nu are loc nici o sosire. Timpul acesta t este o variabilă aleatoare descrisă de o lege exponentială

$$p(t) = e^{-\lambda t}$$

Cu aceste legi se pot calcula medii, dispersii etc. pentru variabila aleatoare discretă r care ia numai valori naturale si pentru variabila aleatoare de tip continuu t.

Mecanismul servirilor

Pentru cuprinderea în model a mecanismului servirilor sunt necesare următoarele:

- O descriere a resurselor necesare pentru a realiza servirea
- O distributie a duratelor servirii
- Numărul de posturi de servire disponibile
- Pozitia cozii/cozilor fată de posturile de servire – fiecare post poate avea coada sa sau coada poate fi unică pentru toate posturile
- Dacă este permis tratarea unor urgente – un post poate eventual întrerupe actiunea de servire curentă pentru a servi un alt client care reprezintă o urgentă

Se admite uzual că durata de servire nu depinde în nici un fel de procesul sosirilor. Se admite, de asemenea, frecvent că durata servirii este o variabilă aleatoare distribuită exponenetial.

Caracteristicile cozilor

Cozile sunt caracterizate prin:

- O disciplină a cozii care poate fi de tipul primul-sosit-primul-servit, de tipul ultimul-sosit-primul-servit sau cu servire la întâmplare
- Tipul clientilor care pot renunta la serviciu dacă coada este prea mare, sau care pot renunta dacă timpul de asteptare a depăsit o anumită limită, sau care schimbă coada dacă au credinta că astfel sunt serviti mai repede
- Capacitatea cozilor – poate fi finită sau (practic) infinită

Schimbarea disciplinei în cozi, cu alte cuvinte a regulei după care este ales clientul următor pentru a fi servit poate uneori să reducă congestia, aglomerarea. Adesea disciplina cozii serveste-mai-întâi-cazurile-simple (clientii care necesită un timp de servire mai scurt) produce un timp mai scurt de asteptare a clientilor în coadă.

În continuare sunt tratate mai în detaliu câteva sisteme cu asteptare.

Sistem cu sosiri simple si un unic post de servire

Sistemul este cu sosiri aleatoare si cu durata servirii, de asemenea aleatoare. Numărul mediu de sosiri (rata medie a sosirilor) λ si numărul mediu de serviri (rata medie a servirilor) μ în unitatea de timp se presupun cunoscute. Cum s-a spus, aceste numere determină procesul aleator în cadrul căruia interactionează clientii cu postul de servire. Sosirile sunt poissoniene de parametru λ, $1/\mu$ este timpul mediu de servire, media unor durate de servire aleatoare care se distribuie conform unei legi exponentiale de parametru μ.

În intervalul $(t, t+\Delta t)$ se poate produce sau nu o sosire cu probabilitătile aproximative $\lambda\,\Delta t$, respectiv $1 - \lambda\,\Delta t$. Probabilitătile ca un client să fie servit (în sensul încheierii actiunii de servire) în acelasi interval sunt, pe acelasi principiu, $\mu\,\Delta t$, respectiv $1 - \mu\,\Delta t$.

Starea postului de servire la un moment dat poate fi *ocupat* (O) sau *liber* (L). Evolutia sistemului într-un interval de timp scurt, Δt este descrisă de următoarele evenimente si tranzitii cu probabilitătile lor:

1. Sistemul este liber si nu se produce nici o solicitare/sosire nouă;
2. Sistemul este ocupat si în intervalul respectiv are loc o servire;
3. Sistemul este liber si apare un client;
4. Sistemul este ocupat si nu are loc încheierea nici unei serviri.

Primele două situatii aduc sau mentin sistemul în starea L, asadar probabilitatea ca la sfârsitul intervalului Δt sistemul să fie în starea L este

$$p_L(t+\Delta t) = p_L(t)(1-\lambda\Delta t) + p_O(t)(\mu\Delta t)$$

Ultimele două situatii aduc sau mentin sistemul în starea O si probabilitatea ca la finele intervalului Δt sistemul să fie în starea O este

$$p_O(t+\Delta t) = p_L(t)(\lambda\Delta t) + p_O(t)(1-\mu\Delta t)$$

În relatiile acestea se multiplică probabilităti atunci când este vorba de intersectia a două evenimente independente ("sistemul-este-liber" si "nu-se-produce-nici-o-sosire", de pildă), se adună probabilităti atunci când se iau în considerare reuniuni de evenimente mutual incompatibile [de pildă, ("sistemul-este-liber" si "nu-se-produce-nici-o-sosire") sau ("sistemul-este-ocupat" si "se-produce-încheierea-unei-serviri"), cu evenimentele compuse dintre paranteze reciproc exclusive].

Sistemul este considerat a fi fără asteptare: un client care soseste si găseste sistemul ocupat poate găsi serviciul căutat în altă parte, deci părăseste imediat sistemul.

Dacă se trece la limită, $\Delta t \to 0$, atunci cele doua ecuatii cu diferente de mai sus pot fi scrise ca un sistem de ecuatii diferentiale

$$\frac{d}{dt}p_L(t) = -\lambda p_L(t) + \mu p_O(t)$$

$$\frac{d}{dt}p_O(t) = -\mu p_O(t) + \lambda p_L(t)$$

care descrie evolutia sistemului ca proces aleator. Dacă se tine seamă de faptul că cele două stări sunt mutual exclusive si una contrara celeilalte, atunci $p_L(t) + p_O(t) = 1$, si sistemul de ecuatii diferentiale se reduce la una singură, pentru a face o alegere

$$\frac{d}{dt} p_L(t) = -\lambda p_L(t) + \mu[1 - p_L(t)]$$

În regim stationar, regimul atins după un timp îndelungat când derivatele temporale se anulează, ecuatiile de mai sus produc sistemul algebric

$$0 = -\lambda p_L(t) + \mu p_O(t)$$
$$0 = -\mu p_O(t) + \lambda p_L(t)$$

cu solutia

$$p_L(\infty) = \frac{\mu}{\lambda + \mu} \qquad p_O(\infty) = \frac{\lambda}{\lambda + \mu}$$

care se mai poate exprima si în functie de asa-numitul grad de încărcare al sistemului $\rho = \lambda / \mu$ sub forma

$$p_L(\infty) = \frac{1}{1 + \rho} \qquad p_O(\infty) = \frac{\rho}{1 + \rho}$$

Se poate face acum o discutie relativ la modul cum depind cele două probabilităti de valorile parametrului $\rho \in (0, +\infty)$. O încărcare foarte redusă, ρ apropiat de zero prin raritatea clientilor (λ mic) sau/si printr-un ritm foarte sustinut al servirilor (μ mare), face ca sistemul să fie mai ales liber, $P_L(\infty) = 1$; o încărcare foarte mare, ρ mare prin solicitări foarte frecvente din partea clientilor (λ mare) sau/si printr-un ritm foarte lent al servirilor (μ mic), face ca sistemul să fie mai ales ocupat, $P_O(\infty) = 1$.

Sistem cu sosiri multiple si nelimitate si o singură statie de servire

Un asemenea sistem beneficiază de un număr infinit de clienti. Sistemul cunoaste mai multe stări: o stare corespunde situatiei cu nici un client în sistem, alte stări sunt cu 1 client în sistem, care este si în curs de servire, cu 2, cu 3 etc. clienti în sistem, unul în curs de servire, ceilalti în asteptarea serviciului căutat. Se presupune că nu există alternativă, nu există un alt loc care să ofere acelasi serviciu, asadar clientii se înscriu în coadă si asteaptă să fie serviti.

Situatiile posibile pentru un interval de timp scurt Δt sunt următoarele:

1. n clienti în asteptare, nici o sosire nouă, nici un client servit, cu probabilitătile multiplicate
$$p_n(t)(1 - \lambda\Delta t)(1 - \mu\Delta t)$$
2. $n - 1$ clienti în asteptare, o sosire, nici un client servit, cu probabilitătile, tot asa, multiplicate

$$p_{n-1}(t)(\lambda\Delta t)(1-\mu\Delta t)$$

3. $n+1$ clienti în asteptare, nici o sosire, un client servit care părăseste sistemul, cu probabilitătile multiplicate si de data aceasta

$$p_{n+1}(t)(1-\lambda\Delta t)(\mu\Delta t)$$

A patra situatie nu există sau, mai larg spus, alte situatii sunt practic excluse, cum ar fi de pildă două sau mai multe sosiri si/sau serviri în intervalul scurt Δt. Asadar, cele trei situatii enumerate sunt mutual exclusive si alcătuiesc un sistem complet de evenimente. Ele sunt premisa tranzitiei după Δt către starea cu n clienti în sistem. Se poate prin urmare scrie

$$p_n(t+\Delta t) = p_{n-1}(t)(\lambda\Delta t)(1-\mu\Delta t) +$$
$$+ p_n(t)(1-\lambda\Delta t)(1-\mu\Delta t) + p_{n+1}(t)(1-\lambda\Delta t)(\mu\Delta t)$$

care exprimă tocmai probabilitatea ca după intervalul Δt sistemul să se afle în starea n. Trecerea la limită, $\Delta t \to 0$, si scrierea ecuatiei pentru toate valorile n posibile generează un sistem de ecuatii diferentiale

$$\frac{d}{dt}p_n(t) = \lambda p_{n-1}(t) - (\lambda+\mu)p_n(t) + \mu p_{n+1}(t) \quad (n = 1,2,...)$$

la care se adaugă varianta specială pentru $n = 0$

$$\frac{d}{dt}p_0(t) = -\lambda p_0(t) + \mu p_1(t)$$

Desigur, regimul dinamic este interesant prin el însusi. El descrie evolutia sistemului pornind de la o stare dată, până atinge după un timp îndelungat un echilibru, o stare stationară. Starea stationară către care tinde sistemul este descrisă de probabilitătile

$$p_n = (1-\rho)\rho^n \quad (n = 0,1,2,...)$$

cu ρ acelasi factor de încărcare sau de utilizare, raportul între ritmul mediu a sosirilor λ si ritmul mediu al servirilor μ.

Probabilitătile stationare trebuie să fie numere pozitive si subunitare. Aceasta se întâmplă numai dacă $\rho < 1$. Calculul lungimii medii a firului de asteptare utilizează formula binecunoscută

$$m = \sum_{n=0}^{\infty} np_n = \frac{\rho}{1-\rho}$$

Se observă că pe măsură ce factorul de încărcare creste si se apropie de unitate, lungimea medie a sirului de clienti în asteptare creste. La limită, $\rho \to 1$, lungimea cozii tinde către ∞, adică sistemul devine instabil. Pentru o încărcare dată se poate calcula si o dispersie a lungimii sirului de asteptare. Aceasta este

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

O mărime importantă sub aspect economic este timpul mediu de asteptare a clientului generic. Expresia acestui timp este dat de relatia de mai jos, alături de care este dată si formula pentru timpul mediu total de asteptare.

$$w = \frac{1}{\lambda}\frac{\rho}{1-\rho}, \qquad mw = \frac{1}{\lambda}\left(\frac{\rho}{1-\rho}\right)^2$$

Ce se întâplă dacă ritmul sosirilor este mai mare decât ritmul servirilor? Evident, sunt necesare mai multe posturi de servire. De aceea sunt tratate în continuare sisteme mai complexe.

Sisteme cu mai multe posturi de servire în paralel

Coada pentru un astfel de sistem se presupune a fi unică. Clientii râmân si asteaptă să fie serviti.
Dacă ritmul mediu al sosirilor este λ clienti în unitatea de timp si timpul mediu de servire a unui client este $1/\mu$, atunci ritmul mediu de servire este variabil cu numărul de clienti prezenti în sistem. Dacă numărul de posturi este m atunci ritmul mediu al servirilor pentru situatiile când un singur client este în sistem este $\mu_1 = \mu$, dacă sunt doi clienti în sistem ritmul mediu al servirii este $\mu_2 = 2\mu$, si pentru $k \le m$ clienti în sistem ritmul mediu de servire este $\mu_k = k\mu$. Numai dacă clientii în sistem sunt în număr mai mare decât numărul posturilor de servire are loc o plafonare a ritmului mediu de servire la $m\mu$ clienti în unitatea de timp si se formează un fir de asteptare.
Într-un interval de timp scurt Δt, probabilitatea sosirii unui client în sistem este, s-a mai spus, aproximativ egală cu $\lambda\Delta t$, iar probabilitatea servirii unui client este, functie de numărul de posturi ocupate, circa $\mu_k\Delta t$. Dacă probabilitatea lipsei de clienti în sistem este notată cu p_0 atunci probabilitătile stationare ale celorlalte stări, cu 1, cu 2, cu 3 s.a.m.d. clienti în sistem se pot calcula din aproape în aproape cu relatiile

$$p_1 = \frac{\lambda}{\mu}p_0, \; p_2 = \frac{\lambda}{2\mu}p_1 = \frac{\lambda^2}{2!\mu^2}p_0, \; p_k = \frac{\lambda}{k\mu}p_{k-1} = \frac{\lambda^k}{k!\mu^k}p_0$$

atât timp cât $k \le m$.
Ecuatia cu diferente

$$p_k(t+\Delta t) = p_{k-1}(t)(\lambda\Delta t)(1-\mu_{k-1}\Delta t) +$$
$$+ p_k(t)(1-\lambda\Delta t)(1-\mu_k\Delta t) + p_{k+1}(t)(1-\lambda\Delta t)(\mu_{k+1}\Delta t)$$

scrisă după aceleasi reguli ca si ecuatia din cazul sistemului cu post de servire unic, prin trecere la limită pentru Δt, produce ecuatia diferentială

$$\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + \mu_k)p_k(t) + \mu_{k+1}p_{k+1}(t)$$

Aceasta, scrisă pentru toti indicii $k = 1,2,\ldots$, tinând seama de relatia cu numărul total de posturi de servire, conduce la relatiile de calcul al probabilitătilor stationare date mai sus, la care se adaugă probabilitătile stărilor cu $k > m$ clienti în sistem

$$p_{m+1} = \frac{\lambda^m}{m!\mu^m}\frac{\lambda}{m\mu}p_0 = \frac{\lambda}{m\mu}p_m$$

si, în general,

$$p_{m+i} = \frac{\lambda^i}{m^i \mu^i}p_m$$

pentru $i > 1$.

Probabilitatea stationară p_0 a stării cu toate posturile de servire libere (nici un client în sistem) se calculează din conditia

$$\sum_{k=0}^{\infty} p_k = 1$$

care este o ecuatie cu necunoscuta p_0

$$p_0\left[1 + \frac{\lambda}{1!\mu} + \frac{\lambda^2}{2!\mu^2} + ... + \frac{\lambda^{m-1}}{(m-1)!\mu^{m-1}} + \frac{\lambda^{m-1}}{(m-1)!\mu^{m-1}}\sum_{i=0}^{\infty}\frac{\lambda^i}{(m\mu)^i}\right] = 1$$

Cu notatia $\rho = \lambda/(m\mu)$ pentru factorul de încărcare, în conditiile $\rho < 1$, suma din expresia de mai sus este suma unei serii geometrice cu ratia subunitară, care are primul termen unitar. Suma respectivă este asadar $1/(1-\rho)$. Continutul parantezei drepte din relatia-ecuatie de mai sus este un număr finit si pozitiv, mai mare decât unitatea. Rezultă că valoarea pentru p_0 are toate caracterisiticile unei probabilităti: este pozitivă si subunitară. Stabilitatea sistemului este posibilă numai dacă rata de utilizare/încărcare este subunitară.

Siruri cu asteptare în serie

Dacă clientii solicită mai multe servicii într-o ordine tehnologică precisă atunci acestia trebuie sa parcurgă un sistem cu servire succesivă, sau serie. Dacă, de pildă, sistemul asigură unor clienti care sosesc într-un ritm mediu de λ în unitatea de timp, două servicii serie cu ratele de servire medie μ_1 si μ_2 serviri pe unitatea de timp atunci starea sistemului este descrisă de două variabile de stare: n_1 numărul clientilor în asteptare si în servire la postul 1 si n_2 numărul clientilor în asteptare si în servire la postul 2.

O enumerare a tranzitiilor si a evenimentelor posibile într-un interval scurt Δt are următoarea înfătisare:

1. $n_1 = 0; n_2 = 0$
   a) nici o sosire în sistem, după Δt $n_1 = 0; n_2 = 0$;
   b) o sosire în sistem, după Δt $n_1 = 1; n_2 = 0$.
2. $n_1 = n_1^0 > 0; n_2 = 0$
   a) nici o sosire, nici o servire, după Δt $n_1 = n_1^0; n_2 = 0$;
   b) nici o sosire, o servire la postul 1, după Δt $n_1 = n_1^0 - 1; n_2 = 1$;

c) o sosire în sistem, nici o servire, după Δt $n_1 = n_1^0 + 1;\ n_2 = 0$.

3. $n_1 = 0;\ n_2 = n_2^0 > 0$

a) nici o sosire, nici o servire, după Δt $n_1 = 0;\ n_2 = n_2^0$;

b) o sosire în sistem, nici o servire, după Δt $n_1 = 1;\ n_2 = n_2^0$;

c) nici o sosire în sistem, o servire la postul 2, după Δt $n_1 = 0;\ n_2 = n_2^0 - 1$.

4. $n_1 = n_1^0 > 0;\ n_2 = n_2^0 > 0$

a) nici o sosire, nici o servire, după Δt $n_1 = n_1^0;\ n_2 = n_2^0$;

b) o sosire în sistem, nici o servire, după Δt $n_1 = n_1^0 + 1;\ n_2 = n_2^0$;

c) nici o sosire, o servire la postul 1, după Δt $n_1 = n_1^0 - 1;\ n_2 = n_2^0 + 1$;

d) nici o sosire, o servire la postul 2, după Δt $n_1 = n_1^0;\ n_2 = n_2^0 - 1$.

S-au omis cu bună stiintă evenimentele constând în producerea concomitentă a unei sosiri si a unei serviri sau a două serviri, din cauza raritătii lor relativ la intervalul scurt Δt luat în considerare.

Ecuatiile-model ale sistemului se scriu în maniera obisnuită. În general

$$\begin{aligned} p(n_1, n_2, t + \Delta t) &\approx (1 - \lambda\Delta t)(1 - \mu_1 \Delta t)(1 - \mu_2 \Delta t) p(n_1, n_2, t) + \\ &+ (\lambda\Delta t)(1 - \mu_1 \Delta t)(1 - \mu_2 \Delta t) p(n_1 - 1, n_2, t) + \\ &+ (1 - \lambda\Delta t)(\mu_1 \Delta t)(1 - \mu_2 \Delta t) p(n_1 + 1, n_2 - 1, t) + \\ &+ (1 - \lambda\Delta t)(1 - \mu_1 \Delta t)(\mu_2 \Delta t) p(n_1, n_2 + 1, t) \end{aligned}$$

Trecerea la limită, $\Delta t \to 0$, produce ecuatia diferentială

$$\begin{aligned} \frac{d}{dt} p(n_1, n_2, t) &= -(\lambda + \mu_1 + \mu_2) p(n_1, n_2, t) + \lambda p(n_1 - 1, n_2, t) \\ &+ \mu_1 p(n_1 + 1, n_2 - 1, t) + \mu_2 p(n_1, n_2 + 1, t) \end{aligned}$$

care trebuie particularizată corespunzător pentru primele trei cazuri enumerate mai devreme

$$\frac{d}{dt} p(0,0,t) = -\lambda p(0,0,t) + \mu_2 p(0,1,t)$$

$$\frac{d}{dt} p(n_1,0,t) = -(\lambda + \mu_1) p(n_1,0,t) + \lambda p(n_1 - 1,0,t) + \mu_2 p(n_1,1,t)$$

$$\frac{d}{dt} p(0,n_2,t) = -(\lambda + \mu_2) p(0,n_2,t) + \mu_1 p(1,n_2 - 1,t) + \mu_2 p(0,n_2 + 1,t)$$

Probabilitătile corespunzătoare regimului stationar sunt date de relatia generală

$$p(n_1, n_2) = \rho_1^{n_1} \rho_2^{n_2} p(0,0)$$

unde s-au notat cu $\rho_1 = \lambda / \mu_1$, respectiv cu $\rho_2 = \lambda / \mu_2$ factorii de încărcare ai sistemului, iar probabilitatea lipsei de clienti în sistem $p(0,0)$ rezultă din conditia

$$\sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \rho_1^{n_1} \rho_2^{n_2} p(0,0) = \frac{p(0,0)}{(1-\rho_1)(1-\rho_2)} = 1$$

adică $p(0,0) = (1-\rho_1)(1-\rho_2)$.

Probabilitǎtile pe cele douǎ siruri de asteptare se calculeazǎ cu relatiile

$$p(n_1) = \rho_1^{n_1}(1-\rho_1) \text{ si } p(n_2) = \rho_2^{n_2}(1-\rho_2)$$

cu interpretarea

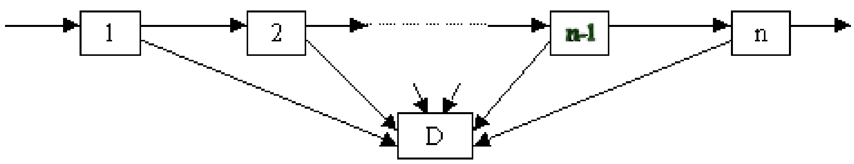
$$p(n_1, n_2) = p(n_1)p(n_2)$$

pentru douǎ siruri independente. O generalizare a acestei proprietǎti la servicii serie cu mai multe (*N*) posturi de servire conduce la relatiile

$$p(n_1, n_2, ..., n_N) = \prod_{i=1}^{N} \rho_i^{n_i}(1-\rho_i)$$

$$p(n_i) = \rho_i^{n_i}(1-\rho_i),\ i = 1,2,...,N$$

Desigur, gama de sisteme de servire a unor clienti solicitatori de servicii este foarte variatǎ. Un caz care meritǎ atentie este cel al sistemelor serie cu depozitare (temporarǎ) alǎturi de firul conditionat tehnologic al serviciilor solicitate si oferite. Schema sistemului este datǎ mai jos si depozitarea este pentru clientii care dintr-un motiv sau altul nu mai pot fi serviti la unul din posturile de servire: fie pentru cǎ servirea precedentǎ a scos din parametrii de calitate produsul sau clientul respectiv, fie din alte motive.
Sub aspect parametric, ritmul mediu al sosirilor la unul din posturile de servire pe firul tehnologic serie diferǎ de la caz la caz, $\lambda_i = \lambda_{i-1} p_{(i-1)i}$, deoarece existǎ posibilitatea ca clientul generic sǎ fie deviat cǎtre depozit cu o probabilitate p_{iD}. Dubla indexare a probabilitǎtilor exprimǎ o tranzitie de la postul $i-1$ la postul *i* sau de la postul *i* la depozitul D. Desigur, $p_{i(i+1)} + p_{iD} = 1$. În aceste conditii, $\lambda_1 = \lambda$ coincide cu ritmul mediu al sosirilor în sistem si $p_{01} = 1$, iar starea stationarǎ a sistemului este descrisǎ de probabilitǎtile asociate lungimilor posibile n_i ale sirurilor de asteptare de la fiecare post de servire, $p(n_i) = \rho_i^{n_i}(1-\rho_i)$ cu $\rho_i = \lambda_i / \mu_i$, rata de încǎrcare a postului *i*, raport între rata medie a sosirilor si rata medie a servirilor la postul respectiv.



Sunt situatii în care capacitatea sistemului este limitatǎ, asadar existǎ o restrictie asupra sosirilor în sistem. În literaturǎ se exemplificǎ cu cazul unui laminor: este posibilǎ o blocare a unitǎtii premergǎtoare ceea ce ar echivala cu riscul rǎcirii pronuntate a prefabricatului care urmeazǎ a fi laminat.
În situatia satisfacerii cererii de un serviciu într-un singur punct de servire se admite cǎ lugimea sirului de asteptare nu trebuie sǎ fie mai mare decât *q*.

Ritmul mediu al sosirilor este λ pentru cazul prezentei în sistem a cel mult q clienti, este nul pentru cazul prezentei în sistem a q clienti.
Ecuatiile care descriu functionarea stationară a sistemului sunt

$$-\lambda p_0 + \mu p_1 = 0$$

$$\lambda p_{n-1} - (\lambda + \mu) p_n + \mu p_{n+1} = 0 \text{ pentru } 1 < n < q$$

$$-\mu p_q + \lambda p_{q-1} = 0$$

Din sistemul de mai sus si din conditia

$$\sum_{n=0}^{q} p_n = 1$$

rezultă

$$p_n = \rho^n \frac{1-\rho}{1-\rho^{q+1}}$$

pentru toti $n \le q$, cu notatiile obisnuite.
Lungimea medie a sirului de asteptare este

$$M(n) = \sum_{n=0}^{q} n p_n = \rho \frac{1-\rho^q(\rho-1)+\rho^{q-1}q}{(1-\rho)(1-\rho^{q-1})}$$

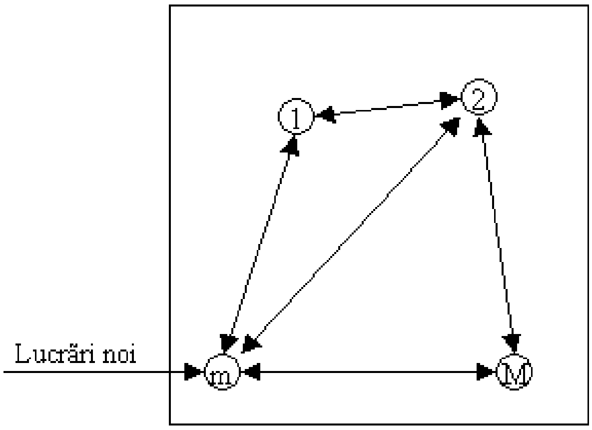
si timpul mediu de asteptare este $M(n)/\lambda$.
Asemănător se pot trata si alte structuri de sisteme cu asteptare, structuri care pot fi o combinatie de substructuri serie si paralel.

Lucrul în ateliere, asteptarea serie si paralel

Se presupune existenta a *M* masini (sau sectii, unităti de servire). La masina *m*, altfel oarecare, sosesc solicitări din două surse, una externă, cealaltă internă, de la celelalte masini. Sosirile la masina *m* sunt aleatoare de ritm mediu λ_m care poate fi diferit pentru masini diferite. Timpii de servire sunt μ_m. La terminarea lucrării la masina *m* are loc un transfer fără consum de timp la altă masină $k, k \ne m$ cu probabilitatea p_{mk}. Este de asemenea probabil ca obiectul servit să părăsească sistemul.
Dacă g_m este frecventa medie de sosire la masina *m*, atunci în regim stationar

$$g_m = \lambda_m + \sum_k q_k p_{km}$$

Se notează cu n_m numărul de clienti de la masina *m*. Dacă punctul *m* nu este o masină ci un centru de masini atunci r_m sunt în servire, $n_m - r_m$ sunt în asteptare. Se notează cu $P_{n_m}(t)$ probabilitatea ca la momentul *t*, la punctul *m* să fie exact n_m clienti. Intrucât sunt *M* masini/centre, se spune că sistemul se află în starea $N = (n_1, n_2, ..., n_M)$ dacă clientii de la masina/punctul $m, m = 1,2,..., M$ sunt în număr de n_m.



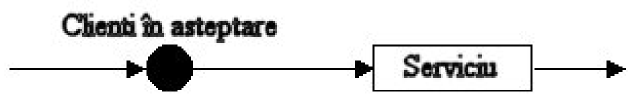
Ecuatia descriptoare este

$$P_N(t+h)=\{1-h\sum\lambda_i-[\sum\alpha_i(n_i)\mu_i]h\}P_N(t)+$$
$$+\sum\alpha_i(n_i+1)\mu_i p_i^* hP_{N+1}(t)+\sum\lambda_i\delta_i hP_{N-1}(t)+$$
$$+\sum\sum\alpha_j(n_j+1)\mu_j p_{ji} hP_{N_2}(t)+o(h)$$

cu notatiile aproape evidente.

SIMULAREA

În cadrul acestui curs si în general, termenul "simulare" are cel putin două semnificatii. Una din semnificatii se leagă de recursul la un model matematic de tip determinist scris pentru un sistem economic si efectuarea de evaluări (repetate) ale comportării si performantelor acelui sistem. Scopul unei astfel de simulări acoperă aspecte de proiectare a sistemelor noi, de ameliorare si de optimizare a sistemelor existente. Semnificatia cealaltă se aplică sistemelor de productie marcate de puternice aspecte stochastice, sisteme în cazul cărora cuprinderea în relatii matematice a tuturor fenomenelor aleatoare guvernante este din punct de vedere practic dificilă sau imposibilă. Această a doua semnificatie face obiectul capitolului prezent.
De notat că "simulare" mai poate însemna si altceva. Se vorbeste, de pildă, de simulatoare de zbor care reproduc comportarea unui avion în zbor dar în realitate simulatorul nu părăseste niciodată solul. Această simulare este utilizată pentru antrenament. Mai există jocuri care simulează o afacere imaginară si jucătorul este în rolul de conducător al afacerii respective cu tinta unor vânzări cât mai bune sau a unui profit cât mai consistent. Acestea sunt simulările unor afaceri.
În sectiunea referitoare la fenomenele de asteptare s-a arătat că, în esentă, un sistem cu asteptare oricât de complex poate fi divizat în subsisteme care constau într-un sir de asteptare si o anumită activitate, asa cum se vede în figura alăturată.



Se poate vorbi, asadar, de subsisteme care servesc *clientii* în asteptare, cu *serviciul* solicitat. Pentru a analiza un asemenea subsistem simplu sunt necesare unele informatii despre procesul de sosire, despre procesul de servire, despre caracteristicile disciplinare ale cozii, despre comportarea clientilor si despre numărul lor (finit sau infinit) etc. Aceste informatii au fost discutate mai în detaliu în capitolul referitor la sistemele cu asteptare. Se reiau aici câteva dintre ele.
Procesul de sosire

- Cum sosesc clientii, pe rând sau în grupuri

- Cum sunt distribuite în timp sosirile, care este distributia statistică a timpilor dintre două sosiri succesive
- Populatia de clienti este finită sau (practic) infinită

Procesul de servire

- O descriere a resurselor necesare ca servirea să înceapă
- Cât durează o servire, o distributie statistică a timpului de servire
- Dacă sunt permise tratări preferentiale (postul de servire ar putea opri servirea unui client pentru a se ocupa de un alt client, de o "urgentă")

Caracteriticile cozii

- Cum sunt alesi pentru a fi serviti clientii, pe principiul FIFO (*first-in first-out*) cunsocut si ca FCFS (*first-come first served* – primul venit primul servit), în maniera LIFO (*last-in first-out*) sau aleator (aceasta se numeste adesea *disciplina cozii*)
- Clientii sunt:
    - Oportunisti, decid să nu rămână la coadă dacă coada este prea lungă
    - Exploratori, părăsesc coada dacă au asteptat deja prea mult
    - Migratori, clientii schimbă cozile dacă ei cred că vor fi serviti mai repede dacă se duc la altă coadă
- Coada este de capacitate finită sau (practic) infinită

De observat că la tot pasul apar situatii de incertitudine, inclusiv relativ la momentul sosirilor si la durata servirilor. Prin urmare, probabilitătile si statistica matematică sunt de neevitat în analiza sistemelor cu asteptare simple sau complexe.

În timp ce teoria sistemelor cu asteptare poate fi utilizată pentru analiza sistemelor simple, sistemele mai complexe afectate de fenomene de asteptare sunt analizate mai curând pe calea simulării, denumită uneori mai precis "simularea sistemelor cu evenimente discrete".

În modelarea sistemelor de productie se procedează frecvent la simularea sistemelor cu evenimente discrete.

Un exemplu de simulare

Ilustrativ pentru simularea evenimentelor discrete este exemplul simplu din figura de mai sus alcătuit dintr-un singur punct de servire si o singură coadă.

Se presupune că clientii sosesc la intervale de timp care sunt distribuite uniform între 1 si 3 minute, adică intervalele de timp dintre două sosiri succesive pot lua orice valoare între 1 si 3 minute si valorile din acest interval sunt egal probabile. Se admite, de asemenea, că durata servirilor este tot o variabilă aleatoare uniform distribuită pe intervalul 0,5 – 2 minute. Iată acum cum se analizează acest sistem simplu prin simulare.

Conceptual, există aici două distributii statistice independente. Gândul simulării poate duce la construirea a două liste lungi de numere – una cu valori ale timpului scurs între două sosiri succesive extrase aleator conform legii de repartitie uniforme pe intevalul (1, 3), cealaltă cu duratele servirilor succesive extrase aleator de un generator de numere aleatoare uniform repartizate în intervalul (0,5, 2). Limbajele de programare de nivel înalt ca si unele pachete de programe gen ***Matlab*** sau ***Excel*** dispun de facilităti variate de generare a numerelor aleatoare. De exemplu în ***Excel***, instructiunile 1+(3 – 1)*RAND() si 0.5 + (2 – 0.5)*RAND() sunt capabile a genera listele de valori necesare în simularea sistemului în studiu. Fie listele de valori din tabelul care urmează:

Intervale între sosiri	Duratele servirii
1,9	1,7
1,3	1,8
1,1	1,5
1,0	0,9
…	…

Pentru usurarea calculelor s-a retinut numai o cifră zecimală din mai multe posibile.
Se admite o stare initială (T = 0) cu nici un client în sistem. Se consultă lista si se formulează întrebarea: "Ce urmează?".
Răspuns: după 1,9 minute apare un (prim) client. Coada este vidă, postul de servire este în asteptare (liber) si clientul intră imediat în procedura de servire.
Din nou: "Ce urmează (a se întâmpla)?"
Răspuns: după alte 1,3 minute, adică la T = 1,9 + 1,3 = 3,2, apare clientul următor. Deoarece postul de servire este ocupat clientul intră în asteptare.
"Ce (eveniment/evenimente) urmează?"
Răspuns: la momentul T = 1,9 + 1,7 = 3,6 clientul în servire va fi gata servit si va părăsi sistemul. În acel moment clientul în asteptare intră în servire, servire care se va finaliza la T = 3,6 + 1,3 = 5,4. "Ce urmează?"
Răspuns: după 1,1 minute de la sosirea clientului anterior, adică la T = 3,2 + 1,1 = 4,3, apare clientul următor. Acest nou client se înscrie în coadă deoarece postul de servire este ocupat. "Ce urmează?"
Răspuns: după 1 minut, adică la T = 4,3 + 1,1 = 5,3, apare clientul următor. Acest client se înscrie în coadă – există deja cineva în coadă – asa că acum coada contine doi solicitatori ai serviciului furnizat de posrtul de servire. "Ce urmează?"
Răspuns: la T = 5,4 clientul în curs de servire va fi gata servit si va părăsi sistemul. La acel moment sunt doi clienti în asteptare si, admitând că disciplina în firul de asteptare este de tipul FIFO, primul client din coadă va intra în procedura de servire (care va consuma 1,5 minute si se va termina, asadar, la T = 5,4 + 1,5 = 6,9). "Ce urmează?"
Răspuns: …… etc si se poate continua în această manieră un număr de pasi oarecare functie de timpul disponibil si de răbdare… Calculele relativ la acest prces sunt făcute mai eficient de un calculator.
O recapitulare a pasilor parcursi arată astfel:

Timpul T	Evenimente
1,9	Apare un client, începe servirea care se va termina la T = 3,6
3,2	Apare un client, se asează în coadă
3,6	Se încheie o servire. Clientul din coadă intră la servire care se va termina la T = 5,4
4,3	Apare un client, se asează în coadă
5,3	Apare un client, se asează în coadă
5,4	Se încheie o servire. Primul client din coadă intră la servire care durează până la T = 6,9
etc.	

Se poate observa din secventa de mai sus cum se *simulează*, cum se *reproduce artificial* functionarea sistemului. Simularea, cum s-a ilustrat mai sus, este denumită mai precis *simulare de evenimente discrete* deoarece sunt urmărite în timp evenimente (sosiri de clienti, încheieri de serviri). Sunt interesante numai punctele de pa axa timpului T=1,9; 3,2; 3,6; 4,3; 5,3; 5,4 etc.

Simularea, odată oprită la un număr (mare) arbitrar de pasi poate servi la a calcula parametri statistici variati relativ la sistem. De pildă, se poate evalua media timpului petrecut în sistem (în asteptare si în servire). Cei doi clienti care au fost serviti au petrecut 1,7 respectiv 2,2 minute în sistem (evaluări făcute prin luarea diferentei momentelor plecării din si sosirii în sistem). Pe baza unui număr mai curând modest de observatii, timpul mediu se estimează a fi (1,7 + 2,2)/2 = 1,95 minute.

Se poate face o statistică a lungimii cozii, cum ar fi lungimea medie a cozii. Dimensiunea cozii este 0 pentru $T < 3,2$, este 1 pentru $3,2 < T < 3,6$, este din nou 0 pentru $3,6 < T < 4,3$, este 1 pentru $4,3 < T < 5,3$, este 2 pentru $5,3 < T < 5,4$, asa încât media ponderată cu timpul este

$$[0(3,2 - 0)+1(3,6 - 3,2)+0(4,3 - 3,6)+1(5,3 - 4,3)+2(5,4 - 5,3)]/5,4 =$$
$$= 0,296$$

Este de comentat aici starea sistemului la începutul calculelor: sistemul este fără clienti, este vid. Alegerea acestei stări de pornire, una din mai multe posibile, poate produce rezulate eronate în ceea ce priveste valorile calculate si de aceea este o practică uzuală ca la simulare să se astepte ceva timp până sistemul “se umple”, până când sistemul intră în regim si numai apoi începe colectarea date pentru calculul parametrilor statistici.

Discutie. În simulare, teoria probabilitătilor si statistica joacă un rol atât în datele de intrare cât si în rezultatele pe care simularea le generează. De exemplu, în simularea fluxului de clienti prin casele unui supermarket, date de intare precum numărul de cumpărători prelucrati este reprezentat prin distributii statistice. Rezultatele de genul timpul de asteptare al clientului, lungimea cozilor etc. sunt reprezentate tot de repatitii statistice. În exemplul de mai devreme s-a făcut apel la distributii statistice uniforme.

Sunt câteva probleme de comentat despre simulare.

Tipic, modelul de simulare trebuie executat pe calculator pentru un timp apreciabil pentru ca rezultatele să fie semnificative statistic si de aceea poate fi costisitor sub aspectul timpului cât calculatorul este ocupat.
Rezultatele simulării pe model tind să devină puternic corelate ceea ce înseamnă că estimările evaluate pe baza acestor modele pot fi înselătoare. Corelatia este un termen statistic care înseamnă că două (sau mai multe) variabile sunt dependente una de alta într-o anume manieră descrisă în capitolul dedicat probabilitătilor si statisiticii matematice. Adesea, anumite tehnici de reducere a variantelor pot fi utile pentru a spori exactitatea cu care se fac estimările obtinute din simulare.
În eventualitatea că se modelează un sistem existent pot apărea dificultăti în a valida modelul (sau programul de calcul) pentru a avea siguranta că modelul reprezintă realitatea.
Dacă modelul de simulare este foarte complex atunci este dificil a izola si a întelege ce se întâmplă în model si a deduce relatiile cauză-efect.
Odată în posesia unui model adecvat, acesta poate fi utilizat în mai multe directii.

- Întelegerea functionării curente a sistemului, elaborarea de explicatii coerente ale comportamentului observat. De exemplu, dacă se observă întârzieri inacceptabile în productia unei sectiuni productive, se pune întrebarea "de ce?", "ce factori cotribuie la aceste întârzieri?"
- Explorarea extinderilor sau schimbărilor posibile ale sistemului, de obicei pentru a încerca a-l îmbunătăti. De exemplu, pentru a spori productia fabricii sunt necesare masini suplimentare? Este posibilă o accelerare a lucrului pe masinile existente? Printr-o întretinere a masinilor se poate mări factorul de utilizare? Calificarea personalului este cea potrivită? Care din acesti factori (sau combinatie de factori) ar fi alegerea cea mai bună pentru a creste productia? De observat că uneori o schimbare care reduce congestionarea într-un punct poate fi însotită de o crestere a ei în alt punct. Astfel, trebuie avut în minte acest fapt atunci când se examinează propunerile de schimbare.
- Proiectarea unui nou sistem de la zero sau încercarea de a reproiecta sistemul pentru a satisface (uneori statistic) anumite cerinte la cost minim. De pildă, în (re)proiectarea unui terminal de pasageri dintr-un aeroport, ce niveluri ale resurselor (vamă, posturi de verificare, facilităti pentru bagaje etc.) sunt necesare si cum trebuie amplasate aceste resurse într-un perimetru nou sau într-unul existent.

Simularea a început a fi aplicată la situatii manageriale în anii 50 târzii ai secolului trecut pentru a examina probleme de asteptare si de stocare. Simularea Monte-Carlo a fost utilizată pentru a modela activitătile legate de facilităti cum sunt depozitele de mărfuri sau rezervoarele cu produse petroliere. Problemele de asteptare (de pildă iesirile din supermarket-uri) sunt, de asemenea, printre cele simulate prin metodele Monte-Carlo. Sintagma Monte-Carlo vine de la orasul cu acelasi nume renumit pentru organizarea de jocuri de noroc. Ca si la

jocul de ruletă, foarte popular printre cei care frecventează cazinourile si în operatiile de simulare se obtin numere aleatoare, dar nu prin rotirea unei rulete ci prin generare cu calculatorul.
Avantajele simulării, nu numai decât în opozitie cu teoria cozilor ci mai curând ca o metodă complemetară, sunt enumerate imediat. Astfel:

- Se pot trata mai direct si mai comod comportamentele dependente de timp
- Matematica asociată cu teoria cozilor este dificilă si este validă numai pentru anumite dsitributii statistice pe când matematica simulării este mult mai accesibilă si poate lucra cu orice distributie statistică
- În unele situatii, este practic imposibil a scrie ecuatiile pe care teoria cozilor o pretinde (de pildă, aspecte de genul schimbării între cozi, vitezele de lucru dependente de coadă etc.)
- Simularea este mult mai usor acceptată si înteleasă de manageri decât teoria cozilor

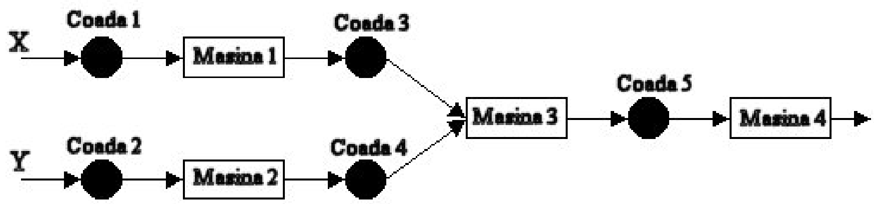
Un dezavantaj al simulării este că este dificil a atinge o solutie optimă cum, de pildă, se obtine rapid si usor cu programarea liniară. O cale de a încerca optimizarea prin simulare ar putea fi alcătuită din etapele următoare:

- Se face o modificare
- Se face o simulare pentru a vedea dacă modificarea aduce sau nu o îmbunătătire
- Se repetă pasii anteriori.

E drept, procesul acesta este consumator de timp-calculator apreciabil.

Un exemplu mai complicat

Fie sistemul descris mai jos, în care sunt două părti X si Y care urmează a fi asamblate, puse laolaltă. Înainte de asamblarea care are loc pe masina 3, atât X cât si Y trebuie să treacă printr-o fază de pregătire pe masina 1, respectiv pe masina 2. După asamblare (X si Y se contopesc într-un ansamblu) mai este necesar un tratament pe masina 4. Evident, dacă o masină este ocupată, părtile sau ansamblul trebuie să astepte procesarea într-o coadă sau alta.



O analiză a situatiei prin simulare se conduce având în vedere existenta celor două componente X si Y si a celor cinci cozi care se pot forma în asteptarea a patru servicii-activităti. Cozile au o disciplină de tipul FIFO – primul-sosit-primul-servit. Aceste elemente si ordinea tehnologică sunt elemente care trebuie clar specificate în orice situatie. O ordine de precedentă arată modul cum “clientii” (X si Y în cazul în discutie) circulă în sistem. Aici X de cum intră în sistem merge mai întâi în coada 1. Coada 1 este urmată de masina 1 si masina 1 la rându-i de coada 3. Coada 3 este urmată de masina 3 si masina 3 este urmată de coada 5 formată înaintea masinii 4.

Regulile de servire spun fiecărui server, fiecărei masini cum să selecteze un client din coada/cozile premergătoare. Regula are semnificatie diferită dacă un server are mai mult de o coadă imediat precedentă. În exemplul în discutie numai masina 3 este în această situatie: coada 3 contine X-i, coada 4 contine Y-i. Pentru a executa servirea la postul “masina 3” este necesară prezenta a cel putin unui X si a cel putin unui Y.

Pentru fiecare din cozi sunt specificate regulile disciplinare. În cazul exemplificat toate regulile sunt de tipul FIFO (first-in-first-out) adică clientii sunt serviti în ordinea sosirii.

Se mai definesc în mod necesar capacitătile cozilor. Dacă o coadă este plină si există o activitatea precedentă, atunci postul de servire care precede acea coadă nu poate elibera un client procesat până când nu se crează loc în coada receptoare.

Pentru clienti (aici, X si Y) este necesară specificarea unei distributii a timpului între sosiri. S-a presupus că pentru părtile X timpii între sosiri sunt repartizati uniform între 0,4 si 0,7 ore. Părtile Y au timpii dintre sosiri distribuiti normal cu media 0,5 si abaterea medie pătratică de 0,2 ore.

Pentru posturile de servire (masinile 1 – 5) este necesar a se specifica distributia duratelor de servire pentru clientii (părtile) pe care îi (le) procesează. Aici masina 1 procesează părtile X într-un timp distribuit normal cu media 0,1 ore si deviatia standard (abaterea medie pătratică) de 0,03 ore. Masina 2 procesează părtile Y cu un timp de servire distribuit normal cu media 0,15 si cu deviatia standard de 0,04 ore. Masina 3 este masina care face asamblarea de X si Y si timpul de procesare este constant, de 0,3 ore.

În final, masina 4 procesează ansambluri de părti si, deoarece nu se poate sti cu sigurantă care din cele două părti (X sau Y) dau numele ansamblului, se specifică uzual o aceeasi distributia a timpilor de prelucrare pentru ambele părti – o distributie normală cu media 0,6 ore si deviatia standard de 0,13 ore. Moduri mai rafinate si, implicit, mai scumpe de implementare a algoritmul de simulare pot evita mai riguros această posibilă confuzie.

Pentru sistemul mai complex adus în discutie, se prezintă rezultatul simulării timp de 100 de ore. Evident, asta nu înseamnă că trebuie consumate 100 de ore reale pentru a obtine rezultatele simulării. Este vorba de un timp el însusi simulat, care în timp-calculator poate fi de câteva secunde sau zeci de secunde. Colectarea datelor pentru evaluările statistice începe după 20 de ore, asadar după un timp considerat suficient pentru “umplerea” sistemului, “asezarea” lui într-un regim considerat stationar.

În aceste conditii, pe intervalul de timp în care se colectează date, de la T = 20 la sfârsitul simulării, sunt 133 de observatii care urmează a fi supuse analizei. Simularea s-a încheiat la un moment T = 101.11, asadar ceva mai târziu fată de cele 100 de ore propuse pentru simulare. Explicatia acestei depăsiri rezidă în

faptul că pachetul execută calculele de simulare până când întâlneste un eveniment (cum ar fi aparitia unei noi părti componente, un final de servire, etc.) care provoacă o schimbare în sistem si, de aceea, durata simulării este mai mare decât timpul de simulare specificat, $T \geq 100$. Aici, primul eveniment de după $T = 100$, care provoacă o schimbare în sistem se produce la $T = 101.11$.
Rezultatele analizei produse cu un anumit program de calcul arată ca în tabelul care urmează.
Tabelul vorbeste prin numerele pe care le contine. Se poate citi numărul

	X	Y
Numărul de sosiri	148	163
Numărul mediu în sistem	13,95	7,42
Numărul maxim în sistem	21	17
Numărul de unităti ansamblate	133	
Durata medie a procesării	1,15	
Abaterea medie pătratică a duratei de procesare	0,14	
Timpul mediu de asteptare	9,50	
Abaterea medie pătratică a timpului de asteptare	4,47	
Durata medie a trecerii prin sistem	10,55	
Abaterea medie pătratică a duratei trecerii prin sistem	4,46	
Durata maximă a trecerii prin sistem	18,27	

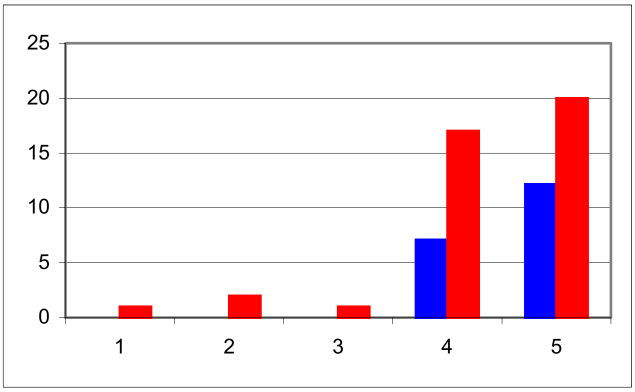
de subansambluri X si Y care au sosit în sistem în intervalul de peste 80 de ore simulate, de la $T = 20$ la $T \approx 100$. Numărul mediu (ponderat cu durate) de unităti X în sistem rezultă a fi 13,95 si, similar, numărul mediu de unităti Y în sistem, de 7,42. S-au finalizat 133 de articole, prin asamblare a câte unui X si a câte unui Y, urmată de procesarea executată pe masina 4. De îndată ce perechile de X si Y devin articole asamblate, care ulterior ies din sistem, valorile calculate sunt unice. Desigur, articolele asamblate pot purta numele X dacă pârtile Y sunt relativ mai putin importante, mai putin voluminoase etc. (similar dacă Y este partea majoră a ansamblului).

Timpul mediu al procesării reprezintă un parametru important. Timpul (aleator) consumat cu procesarea unei părti X (căreia i se adugă la un moment dat o parte Y) pe măsură ce parcurge sistemul este în medie 1,15 ore si are o deviatie standard de cca. 0,14 ore.
La prima vedere ar putea părea de neînteles cum 133 de articole au putut fi terminate dacă ele necesită în medie 1,15 ore fiecare si un total de 133.(1,15) = 152,95 ore. Consumul acesta depăseste cu mult cele 80 de ore pentru care s-a făcut simularea si s-au colectat date. Această posibilă nedumerire se explică prin natura sistemului: procesarea se face simultan pe o bună parte a traseului tehnologic parcurs, adică în paralel pe masini diferite. Luând în considerare graful de mai sus, model al sistemului simulat, la fiecare moment masinile 1, 3 si 4 pot procesa articole X. Astfel, în cele 80 de ore pe parcursul cărora s-au cules datele, timpul disponibil este de cel mult 3.(80) = 240 de ore.
Pentru timpul de asteptare a unui articol în sistem, media este 9,50 si deviatia standard este 4,47 ore. Teoretic, timpul mediu de procesare adunat cu timpul mediu de asteptare (aici 1,15 + 9,50 = 10,65) ar trebuie să coincidă cu timpul mediu de trecere prin sistem (dat mai sus ca fiind 10,55). Aici există o mică diferentă datorată modului cum părtile X si Y, separat. se miscă în sistem.
Din "observatiile" generate prin simulare se pot estima alti parametri cum ar gi gradul de ocupare a posturilor de servire, aici cele patru masini. Iată în tabelul următor aceste estimări.

	Factori de utilizare	Durata medie a procesării	Deviatia standard a duratelor de procesare	Durata de procesare maximă	Clienti procesati
Masina 1	18,1%	0,098	0,033	0,196	148
Masina 2	29,9%	0,147	0,041	0,255	163
Masina 3	55,5%	0,300	0,001	0,300	148
Masina 4	100%	0,603	0,127	0,903	133

Se observă o folosire foarte intensă a masinii 4. Fată de celelalte, masina 4 este de departe cea mai ocupată si, de aceea, aici apare o strangulare în ceea ce priveste capacitatea de productie a sistemului.
Tot pe baza simulării se poate face o analiză statistică a fiecărei cozi. Tabelul următor este o sinteză a unei astfel de analize.
Desigur, numărul de zecimale care apar în unele pozitii din tabel este discutabil. Calculatoarele pot da chiar mai multe cifre după virgulă, dar nu utilitatea lor este subiectul dicutiei curente.
Tinând cont de toate zecimalele sau de mai putine rezultă clar că C5 este coada cea mai importantă, ceea ce confirmă observatia făcută mai devreme, că masina 4 este un punct de strangulare în procesul de productie. Importanta comparativă a cozilor poate fi cuprinsă si într-un grafic. Graficul cozilor din tabel este prezentat imediat.

Coada	Lungimea medie	Lungimea maximă	Asteptarea medie (ore)	Deviatia standard a asteptării	Asteptarea maximă
C1	0	1	0	0	0
C2	0,0115	2	0,0076	0,0434	0,3819
C3	0	1	0	0	0
C4	7,0778	17	3,4622	1,9881	7,5660
C5	12,1586	20	6,6650	2,8512	10,9684



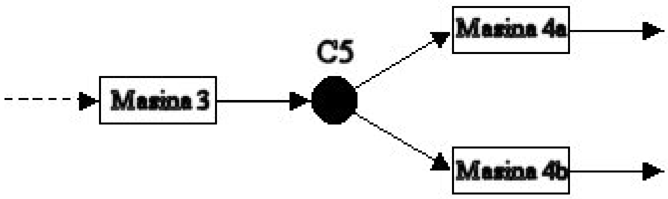
Producerea unor grafice de acest gen este obisnuită în simularea sistemelor cu evenimente discrete. Graficul oferă o imagine a sistemului si a functionării lui uneori mai clară decât un tabel cu numere.

Schimbarea sistemului

Cu un model al unui sistem de productie, cum este modelul de mai sus, obtinut din date de simulare se poate percepe mai complet comportarea acelui sistem. Extinzând discutia la sistemele reale, posibil mai complexe si cu interactiuni mai complicate, fără explorarea numerică oferită de simulare acestea ar fi încă mai dificil de înteles, date fiind variatiile statistice ale duratelor de procesare si ale momentelor de sosire, ale capacitătii cozilor etc.

În cazul sistemului relativ simplu de mai sus, de pildă, s-a evidentiat strangularea de la masina 4 în termeni de volum al productiei. De aici formularea si posibila rezolvare a unei probleme vizând îmbunătătirea sistemului prin explorarea pe cale numerică a unor schimbări ale sistemului. Examinarea datelor simulate îndeamnă la optiunea de a schimba disciplina cozii C5, coada premergătoare masinii 4 unde apare strangularea. Rezultatul poate fi diferit de cel dorit/asteptat. Asa se întâmplă dacă, de pildă, disciplina FIFO (disciplina curentă de tipul primul-venit-primul-servit) este schimbată în MaxWorkDone, care constă în a alege ca articol următor la procesare pe acela

care are deja acumulată o durată maximă de procesare. Rezultatele sunt aproape identice: acelasi număr de articole procesate complet cu timpul mediu de asteptare usor modificat. Articolele stau în medie mai putin timp în sistem desi volumul total al productiei sistemului este neschimbat.
O altă optiune care merită explorată pentru a încerca o sporire a iesirii sistemului este înlocuirea masinii 4: o masină mai productivă sau două masini de acelasi tip în acel punct al sistemului de productie ar trebui să aducă un spor substantial în iesirea sistemului. Dacă productivitatea în punctul sesizat ca fiind îngust se dublează, pare de bun simt a spera ca productia însăsi să se dubleze. Dar: se va dubla productia sau nu, dacă nu se dublează cu cât va creste? Această nouă situatie se poate explora numeric uzând de modelul elaborat corectat cu actiunea diferită a sectiunii ultime a sistemului care arată acum ca în figura următoare.



Coada C5 alimentează acum ambele masini (marcate cu 4a si 4b) în regimul FIFO. Vor trebui acum generate si introduse în calcul durate aleatoare de servire pentru fiecare din cele două masini. Rezultatele evaluărilor prin simularea pentru aceeasi perioadă de timp ca mai sus sunt sintetizate în tabelul care urmează.
Se constată imediat că asteptarea de dublare a productiei a fost iluzorie: productia a crescut de la 133 de articole (aproximativ 133/80 = 1,7 articole pe oră) la numai 151 de articole (aproximativ 151/80 = 1,9 articole pe oră). Aceasta este o crestere cu circa 0,2 articole pe oră, procentual cu circa 12%. Există o explicati pentru asta?
Este firesc a suspecta că durata prea scurtă de simulare a sistemului ar putea deforma rezultatele. Rezultatul după dublarea masinii 4 cu una similară este de 151 de articole în cca. 80 de ore, adică de 1,9 articole pe oră. Ce se întâmplă dacă simularea se extinde pe o durată de 10 ori mai lungă, de pildă pe 1000 de ore, cu începerea colectării de date la ora 200? Rezultatul simulării: productia este în (aproximativ) 800 de ore de 1448 articole, adică este de 1,8 articole pe oră. Este limpede, asadar, si din simularea pe o durată mai îndelungată că efectul angajării în procesul productiv a două masini 4 nu dublează productia. De ce, oare? Răspunsul este relativ simplu: sistemele cu cozi aleatoare si cu activităti de durate aleatoare, cum este cazul aici, sunt notoriu dificile în a fi analizate prin recursul la “bunul simt”. Această situatie este obisnuită în simulare. Pentru a întelege ce se întâmplă, în loc de a folosi intuitia cu toate capcanele ei trebuie examinate pertinent si în detaliu rezultatelor simulării.

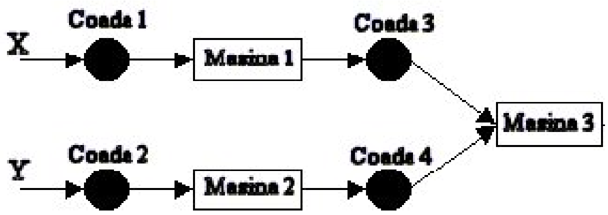
	X	Y
Numărul de sosiri	150	168
Numărul mediu în sistem	1,87	11,55
Numărul maxim în sistem	3	20
Numărul de unităti ansamblate	151	
Durata medie a procesării	1,15	
Abaterea medie pătratică a duratei de procesare	0,13	
Timpul mediu de asteptare	5,45	
Abaterea medie pătratică a timpului de asteptare	1,75	
Durata medie a trecerii prin sistem	6,49	
Abaterea medie pătratică a duratei trecerii prin sistem	4,46	
Durata maximă a trecerii prin sistem	10,13	

Încă două tabele, cu încărcarea masinilor si cu caracteristicile cozilor vor fi suportul altor consideratii si explicatii relativ la functionarea sistemului.

	Factori de utilizare	Durata medie a procesării	Deviatia standard a duratelor de procesare	Durata de procesare maximă	Clienti procesati
Masina 1	19,2%	0,102	0,032	0,196	150
Masina 2	32,1%	0,153	0,042	0,269	168
Masina 3	56,3%	0,300	0,001	0,300	150
Masina 4a	58,6%	0,594	0,122	0,851	79
Masina 4b	53,4%	0,593	0,126	0,958	72

Coada	Lungimea medie	Lungimea maximă	Asteptarea medie (ore)	Deviatia standard a asteptării	Asteptarea maximă
C1	0	1	0	0	0
C2	0,0101	1	0,0048	0,0257	0,2297
C3	0	1	0	0	0
C4	11,2485	20	5,5158	1,7172	8,9479
C5	0,0001	1	0,0001	0,0009	0,0116

Pentru exemplul particular prezentat, cu modificarea structurală operată, răspunsul la mirarea că sistemul nu produce dublu se află în statistica cozilor. Coada C4 are acum de departe cea mai mare lungime medie: 11,2485 în simularea de 100 de ore. Prin examinarea portiunii din graful sistemului reprezentat mai jos, se poate vedea ca în coada C4 părtile Y asteaptă să fie asamblate cu părti X. Din ce cauză coada C4 este lungă?



Coada C4 este lungă fie pentru că masina 3 este deplin utilizată, fie pentru că nu sunt părti X gata pentru a fi asamblate cu părti Y. Din datele despre utilizarea serverelor se vede că masina 3 nu este complet utilizată, astfel că problema este lipsa de părti X în coada C3. Lipsa aceasta ar purtea fi cauzată de o strangulare pe masina 1, dar se vede în tabelul cu încărcarea masinilor că nu este asa, la fel din statistica cozii C1.
În final, se obtine răspunsul la întrebarea, de ce aducând în sistem încă o masină 4 productia nu se dublează: în sistem nu intră suficiente părti X pentru a face posibilă mărirea încă mai importantă a productiei.
Rationamentul sugerat mai sus pare a fi suficient de logic si poate fi un model de utilizare a simulării si în alte împrejurări, chiar mai complexe decât aceasta.

PROGNOZE

Prognoza este estimarea valorilor pe care le poate lua o variabilă sau un set de variabile într-un viitor mai apropiat sau mai îndepărtat. Capitolul prezent prezintă câteva metode de prognozare utilizate în economie.

Uzual, executarea unei operatii de prognozare este premergătoare luării unei/unor decizii, elaborării unor planuri de viitor. Tipic, exercitiul prognozant lucrează în ideea că **dacă se poate anticipa viitorul, atunci comportamentul individual sau al unui agent economic se poate modifica de pe acum, astfel ca la vremea când viitorul va deveni prezent, individul sau agentul economic să se afle într-o pozitie bună, mai bună decât ar putea fi fără o asemenea anticipare**. Aplicatii ale prognozelor se regăsesc în împrejurări variate. Iată câteva exemple:

- Reglarea stocurilor/planificarea productiei – prognozarea cererii pentru un produs face posibil controlul stocurilor de materii prime si de produse finite, planificarea productiei etc.
- Politica de investitii – prognozarea unor informatii financiare de genul evolutiei dobânzilor, a ratelor de schimb, a valoarii actiunilor, a pretului aurului etc. Este vorba aici de un vast domeniu în care nimeni n-a dezvoltat încă metode de prognozare consistente si precise (sau cel putin, aceia care le detin nu le-au spus nimănui!)
- Politica economică – prognozarea informatiilor economice, cum sunt cresterea economică, somajul, rata inflatiei etc. este de importantă vitală în planificarea viitorului, atât pentru guverne cât si pentru mediul de afaceri.

Imaginatia permite, cel putin pentru un moment, a accepta că autorul (sau cititorul) acestei lucrări s-ar afla în fata unei bune prezicătoare a viitorului, de pildă o zână cumsecade, care îi spune că apreciază bunătatea, virtutile si pregătirea exceptională în profesie (e doar un basm, desigur) si a decis să-i furnizeze trei prognosticuri la alegere. Care trei lucruri în viata personală si/sau de om de afaceri ar fi cele mai interesante pentru omul obisnuit? Probabil că, în ordinea descrescândă a importantei, acestea ar fi:

- Data decesului
- Numerele câstigătoare la tragerea următoare a loteriei nationale
- Numerele câstigătoare la tragerea de după aceasta a loteriei nationale

După cum se observă din lista de răspunsuri anticipative propusă, prognozele au legătură cu probleme de viată si de moarte si au consecinte legate de viată si de moarte. Este de asemenea clar că pentru a face anumite previziuni, de pildă asupra datei decesului, trebuie ca, în absenta ajutorului zânei celei bune, să se colecteze ceva date pentru a face posibilă o prognoză mai documentată si, de sperat, mai precisă. De pildă, autorul ca persoană ar trebui să afle speranta de viată a cadrelor didactice (care nu fumează, care beau moderat, care nu practică exercitiile fizice) din universitătile din România. Autorul ar putea să se supună unui temeinic examen medical. Ideea generală este că o colectie de date relevante *poate* duce la o prognoză acceptabilă ca precizie. Se poate însă ca datele acestea să nu ducă la o prognoză prea exactă: autorul ar putea fi călcat de o masină chiar mâine si ar fi transferat "dincolo" mult înainte de termenul prognozat…

Tipuri de probleme si metode de prognozare

O clasificare a problemelor de prognoză tine seamă de scara de timp implicată, adică orizontul de timp în viitor pentru care prognoza se speră a fi acoperitoare si valabilă. Uzual, claselor de prognoze li se asociază calificativele "pe termen scurt", "pe termen mediu" si "pe termen lung". Semnificatia acestor calificative este variabilă în functie de contextul situatiei studiate. O prognoză asupra cererii de energie, care ajută la un program de constructie de centrale este "pe termen scurt" dacă se referă la următorii 5-10 ani, dar este "pe termen lung" dacă se ocupă de următorii 50 de ani. În multe situatii, prognoza pe 6 luni asupra cererii consumatorilor de un anumit produs/serviciu este una "pe termen scurt", dar extinsă la câtiva ani este "pe termen lung". Tabelul următor arată câteva scări de timp asociate cu decizii economice.

Scara de timp	Tipul deciziilor	Exemple
Pe termen scurt (până la 3 - 6 luni)	Operationale	Controlul stocurilor Planificarea productiei, distributia
Pe termen mediu (între 3 - 6 luni si 2 ani)	Tactice	Instalatii si echipamente în leasing Variatia ofertei de fortă de muncă
Pe termen lung (dincolo de 2 ani)	Strategice	Cercetare si dezvoltare Achizitii de active, fuziuni Schimbări de profil al productiei

Ratiunea majoră a clasificării de mai sus este aceea că pentru fiecare situatie se aplică un tip anumit de prognoză. Metoda de prognoză potrivită pentru a estima vânzările de luna viitoare (tipică pentru termen scurt) va fi probabil inadecvată pentru a prognoza vânzările pe următorii cinci ani (termen lung). În afaceri volumul de date utilizat în metodele cantitative mentionate variază de la foarte mare pentru prognozele pe termen scurt, la foarte redus pentru prognozele pe termen lung.
Metodele de prognozare pot fi clasificate în mai multe categorii:

- Metode calitative – în care nu există un model matematic formal, adesea deoarece datele disponibile nu par a fi reprezentative pentru viitor (pe termen lung)
- Metode regresionale – o extensie a regresiei liniare în care o variabilă este considerată a fi corelată liniar cu un număr de alte variabile independente
- Metode cu mai multe ecuatii – când există un număr de variabile dependente care interactionează una cu alta prin mai multe ecuatii (ca în modelele economice)
- Metodele seriilor temporale – unde o singură variabilă se schimbă în timp si valorile ei viitoare sunt dependente de valorile ei trecute.

Mai departe sunt expuse pe rând aceste metode.

Metode calitative

Metodele din această categorie sunt utilizate mai ales atunci când datele din trecut pe care s-ar putea baza prognoza sunt considerate irelevante. Aceste metode sunt utilizate aproape exclusiv pentru prognozele pe termen lung. O tratare a prognozelor din această clasă o furnizează metoda *Delphi*.

Vechii greci aveau o conceptie specială asupra prognozelor si credeau că cele mai indicate persoane de consultat sunt zeii. La oracolul de la Delphi din vechea Grecie întrebărilor li se răspundea printr-un *medium*, o femeie de peste 50 de ani, separată de bărbatul ei si îmbrăcată în rochii de fecioară. Dacă cineva dorea un răspuns la o întrebare trebuia:

- Să îi ofere o prăjitură
- Să ofere un animal pentru sacrificare
- Să se îmbăieze cu medium-ul într-un izvor.

După aceste preliminarii medium-ul se aseza pe un trepied în subsolul templului, mesteca frunze de laur si răspundea la întrebări, de obicei în cuvinte ambigue.

Este asadar legitim a întreba dacă în adâncimea unui subsol undeva, există un functionar guvernamental care mestecă frunze de laur si care este angajat pentru a prognoza crestrea economică, succesul în alegeri etc. Probabil există!

O clipă de reflectie: sunt credibile prognozele făcute în această manieră? Oracolul din Delphi producea prognoze precise sau nu?

O anchetă stiintifică recentă, publicată în *New Scientist* din 1 septembrie 2001, arată că medium-ul delira din cauza inhalării unor hidrocarburi (în particular etilenă) emanate dintr-o fisură geologică aflată sub templu.

În zilele noastre metoda Delphi are o semnificatie diferită. Metoda implică chestionarea unui corp de experti pentru a ajunge la un consens asupra înfătisării viitorului. În subtextul ideii de a apela la experti este *credinta* că vederea lor în viitor este mai bună decât aceea a unor non-experti (cum ar fi oamenii alesi la întâmplare pe stradă). Ce tipuri de experti trebuie alesi pentru o încercare de a face o prognoză pentru 50 de ani?

Într-un studiu Delphi expertii sunt consultati separat pentru a evita o parte din influentele care ar putea rezulta dacă s-ar aduna laolaltă, cum ar fi dominarea dezbaterii de un individ cu personalitate puternică, vederile divergente (dar valide) ale multor altora nefiind exprimate de teama umilirii.
O întrebare tipică ar putea fi “În ce an (dacă se va întâmpla vreodată) este de asteptat ca transportul rapid automatizat să devină obisnuit pentru orasele mari din Europa?”. Răspunsurile sunt puse împreună sub forma unei distributii pe ani (cu comentarii atasate) si sunt eventual retrimise la experti pentru a obtine estimări revizuite. Procesul este repetat până când de obtine un (relativ) *consens*. Este clar că o astfel de metodă are multe deficiente dar nu există o cale mai bună de a avea o imagine asupra viitorului în conditiile în care datele relevante necesare pentru metode mai cantitative lipsesc.
Ca un exemplu, în *Science Journal* din octombrie 1967 s-a publicat un studiu Delphi care încerca o privire în viitor. Au trecut suficient de multi ani de atunci pentru a putea aprecia cât de bună a fost prognoza. S-au formulat atunci multe întrebări despre orizontul de timp în care se va întâmpla ceva anume. În continuare se reproduc câteva răspunsuri. Pentru fiecare întrebare s-a acordat o cvartilă superioră de 75% pentru timpul în care expertii apreciau că acel ceva se va produce.
Pentru tranzitul rapid automat cvartila superioară indica anul 1985, asadar expertii credeau în 1967 că în 1985 în cele mai multe zone urbane tranzitul rapid automatizat va fi larg răspândit. Realizarea unui sistem de acest gen va mai lua multi ani…
Răspândirea largă a masinilor de învătat avea cvartila superioară de 75% situată la 1990, adică 75% din expertii chestionati în 1967 credeau că pe la 1990 masinile de învătat rafinate vor fi la tot pasul. Evident, nu aceasta este situatia azi…
Utilizarea pe scară largă a robotilor avea cvartila superioară stabilită la anul 1995: 75% din expertii întrebati în 1967 credeau că în 1995 robotii vor fi extrem de prezenti. Nici această prognoză nu excelează prin acuratete.
Este clar că cel putin aceste prognoze sunt foarte inexacte. Privind critic toate cele 25 de predictii făcute atunci, mai ales cele legate de viată si societate după 1967, multe sunt vădit imprecise.
Asta aduce în prim plan o problemă cheie: *diferenta dintre prognoză si rezultatul observat în realitate* sau *eroarea de prognozare*.
Cu toate acestea, în 1967 când a fost făcut acest studiu Delphi, nu exista o altă posibilitate care să răspundă la acele întrebări.
În multe privinte problema ridicată relativ la calitatea prognozelor nu este dacă o metodă particulară dă rezultate bune ci dacă metoda selectată este cea mai bună metodă accesibilă. Trebuie folosită cea mai potrivită, cea mai bună metodă de prognozare, chiar dacă se cunoaste istoric că ea nu dă prognoze precise.

Metode regresionale

Problema regresiei liniare este deja cunoscută: o dreaptă de ecuatie

$$Y = a + bX$$

este potrivită pe date, uzual prin metode celor mai mici pătrate. Dacă sunt k variabile independente X_1, X_2, …, X_k atunci se caută o relatie de regresie de forma

$$Y = a + b_1X_1 + b_2X_2 + ... + b_kX_k$$

care este o regresie liniară multiplă. Cunoasterea relatiei de regresie crează premisa de a face prognoze prin introducerea unor valori X_i, $i = 1, 2, …, k$ pentru a produce un Y, o valoare prognozată.

Metode cu mai multe ecuatii

Metodele din această categorie sunt utilizate frecvent în modelarea economică dacă există mai multe variabile care interactionează una cu alta prin ecuatii de o formă bazată uzual pe teoria economică. Teoria economică dă o privire asupra structurii relatiilor între diferite variabile. Relatiile numerice exacte între acele variabile sunt deduse adesea din date observate experimental.

Ca un exemplu, iată un model simplu. Fie X – venitul personal, Y – cheltuielile personale, I – investitiile personale si r – rata dobânzilor practicată de bănci.

Din teoria economică se presupune că cheltuielile sunt o functie liniară de venitul disponibil

$$Y = a_1 + b_1(X - a_1)$$

investitiile sunt functie liniară de rata dobânzilor

$$I = a_2 + b_2r$$

si ecuatia de bilant este

$$X = Y + I \quad \text{(venitul = cheltuieli + investitii)}$$

a_1, a_2, b_1, b_2 sunt constante.

Sunt aici trei ecuatii si patru variabile (X, Y, I, r) si pentru rezolvarea acestor ecuatii uneia dintre variabile trebuie să i se atribuie o valoare. Variabila aleasă se numeste *exogenă* deoarece valoarea ei este decisă în afara sistemului; variabilele rămase sunt numite *endogene* si ele sunt determinate ca solutii ale sistemului de ecuatii. De pildă, în modelul de mai sus rata dobânzilor se poate considera exogenă si se poate urmări cum variază X, Y si I atunci când se modifică r.

De obicei, constantele a_1, a_2, b_1, b_2 nu sunt cunoscute exact si trebuie estimate din date experimentale print-o procedură relativ complexă. Aceste constante sunt diferite pentru grupe de oameni diferite si fac diferente de genul urban/rural, bărbati/femei, căsătoriti/necăsătoriti etc.

Există relatii-model care contin mai multe variabile decât în exemplul de mai sus. Adesea, fiecare din variabile are un indice temporal, ceea ce face posibilă cuprinderea a unor aspecte dinamice.

Metodele bazate pe relatii din econometrie au erori de prezicere mari atunci când sunt utilizate pentru prognoze economice la scară mare, de pildă la scara unei natiuni si pe termen mediu. Cu toate acestea, o prognoză, fie ea si modestă ca acuratete este mai bună decât nici o prognoză si dacă există mai multe metode de prognozare trebuie aleasă aceea care pare a fi cea mai potrivită.

Serii temporale, analiză si metode

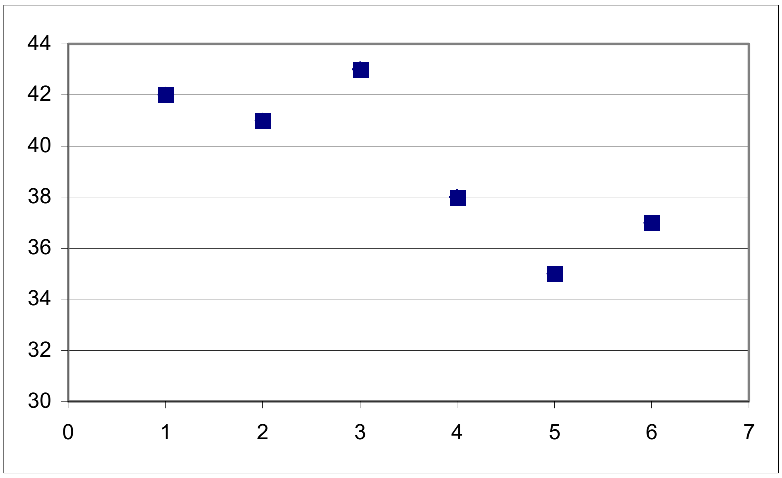
Metodele din această sectiune se aplică variabilelor care se schimbă în timp si despre care se poate spune că depind numai de timp si de valorile anterioare, adică ele nu depind de alti factori externi. Dacă Y_t este valoarea variabilei la momentul t atunci ecuatia pentru Y_t este

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots, Y_0, t)$$

adică valoarea variabilei la timpul t este o functie exclusiv de valorile anterioare si de timp, nici un alt factor extern sau altă variabilă externă nu are vreo relevantă, vreo influentă. Scopul analizei seriilor de timp este de a descoperi natura functiei f si prin aceasta a permite predictia, prognoza pentru variabila Y_t. Metodele legate de seriile de timp sunt eficace mai ales pentru prognozele pe termen scurt unde în limite rationale comportarea trecută a unei anumite varibile este un indicator bun asupra comportării ei în viitorul apropiat. Un exemplu tipic îl constituie prognozarea cererii. Este necesară la acest punct o distinctie între cerere si vânzări: cererea este ceea ce clientii vor, vânzările sunt ceea ce se vinde efectiv si cele două cantităti pot fi diferite.
Datele observate în decursul a sase luni sunt cuprinse în tabelul următor:

Luna	1	2	3	4	5	6
Cererea (x 100)	42	41	43	38	35	37

În reprezentare grafică, relatia $Y_t - t$ este dată mai jos prin puncte.



Analiza care urmează are ca scop a discerne, a evidentia o relatie între valorile Y_t observate până la un moment dat pentru a face posibilă o prognoză asupra valorilor viitoare.

Sunt date imediat în detaliu două tehnici de analiză a seriilor temporale si, mai departe, elementele principale ale unei metode mai rafinate.

Metoda mediei mobile. O metodă foarte simplă de prognozare în cazul seriilor temporale constă în a lua o medie mobilă, uneori o medie mobilă ponderată si a o extinde, a o proiecta în viitor.

Media mobilă m_t relativ la ultimele L perioade observate care se sfârsesc la momentul t se calculează ca media aritmetică a valorilor pentru perioadele $t - L + 1, t - L + 2, t - L + 3, ..., t - 1, t$

$$m_t = (Y_{t-L+1} + Y_{t-L+2} + Y_{t-L+3} + ... + Y_{t-1} + Y_t)/L$$

Pentru a elabora prognoza pentru alte intervale ulterioare lui t, se ia ca valoare prognozată exact m_t. Uzual se prognozează numai o perioadă în viitor si se actualizează media mobilă de îndată ce observatia relativ la perioada imediat următoare lui t devine accesibilă.

Pentru exemplul din tabelul de mai sus care contine cererea de un produs de-a lungul a 6 luni, se poate calcula lunar o medie mobilă pe trei luni si să se prognozeze cererea pe luna a 7-a. Evident, nu se poate calcula o medie (mobilă) pe trei luni până nu s-au acumulat date pentru cel putin trei luni succesive, adică este posibil a se face aceste evaluări numai după ce datele pentru luna a treia sunt cunoscute. Media mobilă la luna a treia este

$$m_3 = (42 + 41 + 43)/3 = 42$$

si mediile mobile pentru lunile următoare sunt

$$m_4 = (41 + 43 + 38)/3 = 40{,}7$$
$$m_5 = (43 + 38 + 35)/3 = 38{,}7$$
$$m_6 = (38 + 35 + 37)/3 = 36{,}7$$

Ca predictie pentru luna a 7-a se utilizează valoarea m_6. Asadar, cererea prognozată pentru luna a 7-a este de 3.670 u.f.

Dar cât de bună este prognoza făcută? Dacă se folosesc medii pe două luni, oare rezultatele nu sunt mai precise? Răspunsul la aceste întrebări se află prin calcul.

Pentru a genera o prognoză asupra cererii din luna a 7-a pe baza mediilor mobile evaluate pe două luni se evaluează mai întâi

$$m_2 = (42 + 41)/2 = 41{,}5$$
$$m_3 = (41 + 43)/2 = 42$$
$$m_4 = (43 + 38)/2 = 40{,}5$$
$$m_5 = (38 + 35)/2 = 36{,}5$$
$$m_6 = (35 + 37)/2 = 36$$

Este o predictie diferită de cea de mai devreme: 3600 u.f. fată de 3670 u.f., rezultatul calculului cu medii mobile pe trei luni. Care din cele două valori este mai de încredere?

Într-o logică simplă, alegerea prognozei celei mai bune se face printr-o interpretare a informatiei disponibile. Astfel, media pe primele trei luni, $m_3 = 42$ se consideră a fi o prognoză pentru luna a patra. Aceasta este prognoza pentru luna a patra. Dar la finele lunei a patra se constată o cerere reală de 38. Se poate calcula o eroare de predicitie

$$eroare = prognoză – observatie = 42 – 38 = 4$$

Eroarea poate fi definită si inversând ordinea termenilor în expresia de mai sus. Se obtin erori cu semn schimbat, valoarea absolută rămânând însă aceeasi (aceasta, de fapt, contează).

În luna a patra se poate face o prognoză pentru luna a cincea $m_4 = 40,7$ dar rezultatul observat în luna a cincea este 35, ceea ce arată o eroare de 40,7 – 35 = 5,7.

În luna a cincea prognoza pentru luna următoare, a sasea, este $m_5 = 38,7$ dar rezultatul efectiv pentru luna a sasea este 37 si eroarea rezultată este 38,7 – 37 = 1,7.

Pe baza acestor rezultate se construieste tabelul următor:

Luna	1	2	3	4	5	6	7
Cererea (x 100)	42	41	43	38	35	37	?
Prognoza	-	-	-	42	40,7	38,7	36,7
Eroare	-	-	-	4	5,7	1,7	?

Dacă se folseste media mobilă pe două luni se poate întocmi un tabel similar:

Luna	1	2	3	4	5	6	7
Cererea (x 100)	42	41	43	38	35	37	?
Prognoza	-	-	41,5	42	40,5	36,5	36
Eroare	-	-	–1,5	4	5,5	–0,5	?

Aceste două tabele sugerează cât de bune sunt prognozele pe cele două căi. Aprecierea se face prin compararea erorilor de predictie evaluate pe datele deja accesibile.

La modul ideal, ar fi de dorit ca toate erorile să fie nule. Asta ar da încredere, poate excesivă încredere, că prognoza pentru luna a saptea este foarte probabil corectă. Dar în realitate erorile nule sunt practic excluse. Este pe de altă parte dificil a privi cele două secvente de numere reprezentând erorile si a le compara. Este mai convenabil si mai eficient a reduce fiecare secventă la o valoare sintetică, usor de obtinut, care să fie o măsură a erorilor, o măsură usor de comparat. O functie potrivită acestui scop este *eroarea medie pătratică.*

Logica cere mai întâi lichidarea deosebirii dintre erorile în plus si erorile în minus: prin ridicare la pătrat toate valorile, pozitive sau negative devin pozitive. Apoi erorile mari au pătrate mai mari, cele mici au valori relativ încă mai mici în urma ridicării la pătrat. O prognoză perfectă ar avea eroarea medie pătratică nulă. Realitatea este diferită de idealitate si în orice împrejurare este de preferat metoda care dă *cea mai mică* eroare medie pătratică.

În exemplul în discutie, dacă de foloseste media mobilă evaluată pe trei luni, eroare pătratică medie este

$$[4^2 + 5,7^2 + 1,7^2]/3 = 17,13$$

si dacă de foloseste pentru media mobilă evaluată pe două luni consecutive, eroare pătratică medie este

$$[(-1,5)^2 + 4^2 + 5,5^2 + (-0,5)^2]/4 = 12,19$$

Cea mai mică dintre aceste două valori este cea din cazul prognozei făcute cu media mobilă pe două luni consecutive si, de aceea, este de preferat metodei celeilalte. În consecintă este retinută pentru luna a saptea prognoza de 3600.

Eroarea medie pătratică este cunoscută si sub denumirea de *deviatia medie pătratică* sau, după extragerea rădăcinii pătrate, sub numele de *deviatie standard*.

De retinut în final faptul că din rationamentul de mai sus rezultă posibilitatea discriminării între două prognoze diferite, una bazată pe media mobilă pe trei luni consecutive, alta pe media mobilă pe două luni la rând. Criteriul este deviatia medie pătratică care trebuie să fie minimă.

O versiune modificată a metodei mediei mobile este *metoda mediei mobile cu ponderi*. Deosebirea fată de original constă în ponderarea diferită a observatiilor grupate în seria temporală de bază, de regulă cu ponderi defavorabile pentru observatiile mai vechi. Această modificare poate fi de multe ori benefică.

La dispozitia celor interesati de prognoze mai există încă circa o duzină de alte metode, multe din ele implementate ca programe de calcul comerciale. În cele de urmează este adusă în prim plan una din acestea.

Netezirea exponentială simplă

Un dezavantaj al utilizării mediei mobile pentru prognozare este faptul că în calcule toate observatiile au ponderi egale, $1/L$, desi este de asteptat ca observatiile mai recente să fie un indicator mai bun a ceea ce va fi în viitor si de aceea, acestea ar trebui să aibă o pondere mai mare. Pe de altă parte, în metoda mediei mobile se folosesc numai observatiile recente. Poate că ar trebui să conteze în oarecare măsură si alte observatii anterioare acestora.

O metodă cunoscută ca *netezirea exponentială* sau, mai precis, *netezirea exponentială simplă* dă o pondere mai mare observatiilor recente si ia în considerare toate observatiile anterioare. Pentru aceasta se precizează o constantă μ pozitivă si subunitară si se calculează o medie mobilă netezită pentru întreaga perioadă t – notată în continuare cu M_t – cu relatia

$$M_t = \mu Y_t + \mu(1-\mu)Y_{t-1} + \mu(1-\mu)^2 Y_{t-2} + \mu(1-\mu)^3 Y_{t-3} + ...$$

Astfel, se iau în cosideratie cu anumite ponderi toate valorile observate, spre deosebire de metoda anterioară care uza numai de o parte din ele.

Relatia de mai sus pare dificilă sub aspectul calculelor dar ea se poate rescrie ca

$$M_t = \mu Y_t + (1-\mu)[\mu Y_{t-1} + \mu(1-\mu)Y_{t-2} + \mu(1-\mu)^2 Y_{t-3} + ...]$$

adică sub forma

$$M_t = \mu Y_t + (1-\mu)M_{t-1}$$

Asadar, media mobilă netezită exponential referitoare la perioada t este o combinatie liniară (convexă) a valorii curente Y_t si a mediei mobile precedente, M_{t-1}, obtinută tot prin netezire exponentială.

Constanta μ este numită *constantă de netezire* si valoarea ei reflectă ponderea atribuită observatiei curente Y_t în evaluarea mediei mobile netezite exponential pentru perioada t, M_t, care este prognoza pentru perioada următoare $t + 1$. De pildă, $\mu = 0{,}2$ arată că ponderea ultimei observatii este de 20%, iar ponderea observatiilor anterioare este de 80%.

O altă scriere a relatiei de mai sus este

$$M_t = M_{t-1} - \mu(M_{t-1} - Y_t)$$

si lectura ei este: prognoza curentă = prognoza anterioară – μ(eroarea în prognoza anterioară) asa încât netezirea exponentială poate fi interpretată ca o prognoză actualizată permanent prin eroarea de predictie cea mai recentă.

Urmează acum un exemplu de calcul pe aceleasi date referitoare la cererea de un anumit produs, utilizate mai sus. Se evaluează succesiv media mobilă netezită exponential cu constanta de netezire $\mu = 0{,}2$. Pentru prima pas, media M_1 se ia totdeauna egală cu Y_1.

$$M_1 = Y_1 = 42$$
$$M_2 = 0{,}2Y_2 + 0{,}8M_1 = 0{,}2(41) + 0{,}8(42) = 41{,}80$$
$$M_3 = 0{,}2Y_3 + 0{,}8M_2 = 0{,}2(43) + 0{,}8(41{,}80) = 42{,}04$$
$$M_4 = 0{,}2Y_4 + 0{,}8M_3 = 0{,}2(38) + 0{,}8(42{,}04) = 41{,}23$$
$$M_5 = 0{,}2Y_5 + 0{,}8M_4 = 0{,}2(35) + 0{,}8(41{,}23) = 39{,}98$$
$$M_6 = 0{,}2Y_6 + 0{,}8M_5 = 0{,}2(37) + 0{,}8(39{,}98) = 39{,}38$$

Numărul de cifre semnificative este o problemă de optiune contextuală: aici este suficient a lucra cu 2-3 cifre după virgulă. Valoarea M_6 este utilizată pentru a prognoza luna a saptea: 3938 u.f.

Dacă se modofică ponderea informatieie proaspete la $\mu = 0{,}9$ se obtin succesiv valorile

$$M_1 = Y_1 = 42$$
$$M_2 = 0{,}9Y_2 + 0{,}1M_1 = 0{,}9(41) + 0{,}1(42) = 41{,}10$$
$$M_3 = 0{,}9Y_3 + 0{,}1M_2 = 0{,}9(43) + 0{,}1(41{,}10) = 42{,}81$$
$$M_4 = 0{,}9Y_4 + 0{,}1M_3 = 0{,}9(38) + 0{,}1(42{,}81) = 38{,}48$$
$$M_5 = 0{,}9Y_5 + 0{,}1M_4 = 0{,}9(35) + 0{,}1(38{,}48) = 35{,}35$$
$$M_6 = 0{,}9Y_6 + 0{,}1M_5 = 0{,}9(37) + 0{,}1(35{,}35) = 36{,}84$$

Ca si mai devreme, M_6 este prognoza pentru luna a saptea, adică 3684 u.f.

Pentru a decide asupra celei mai bune valori pentru μ (între cele două valori 0,2 si 0,9) se calculează valorile pentru eroarea/deviatia medie pătratică (EMP).

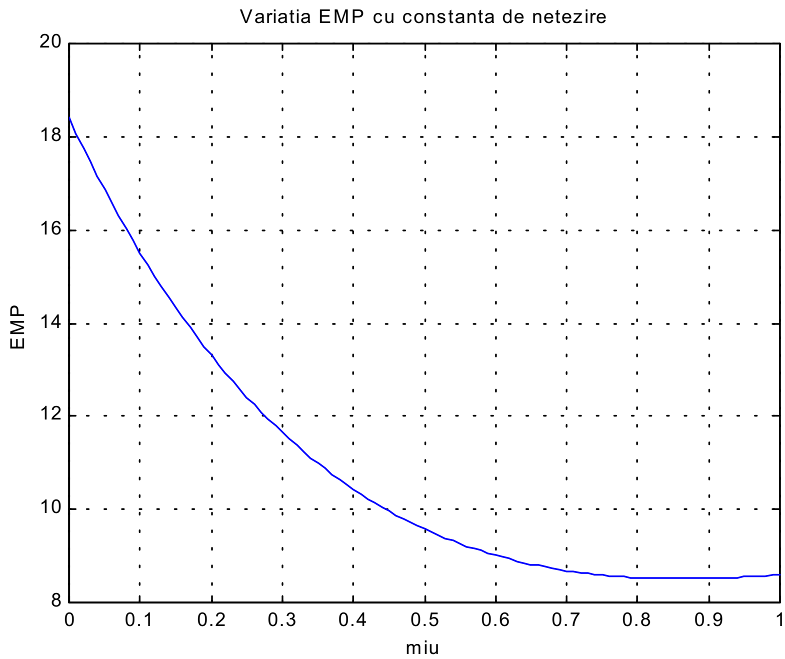
Pentru $\mu = 0{,}2$

$$\text{EMP} = [(42 - 41)^2 + (41{,}80 - 43)^2 + (42{,}04 - 38)^2 + (41{,}23 - 35)^2 + $$
$$+ (39{,}98 - 37)^2]/5 = 13{,}29$$

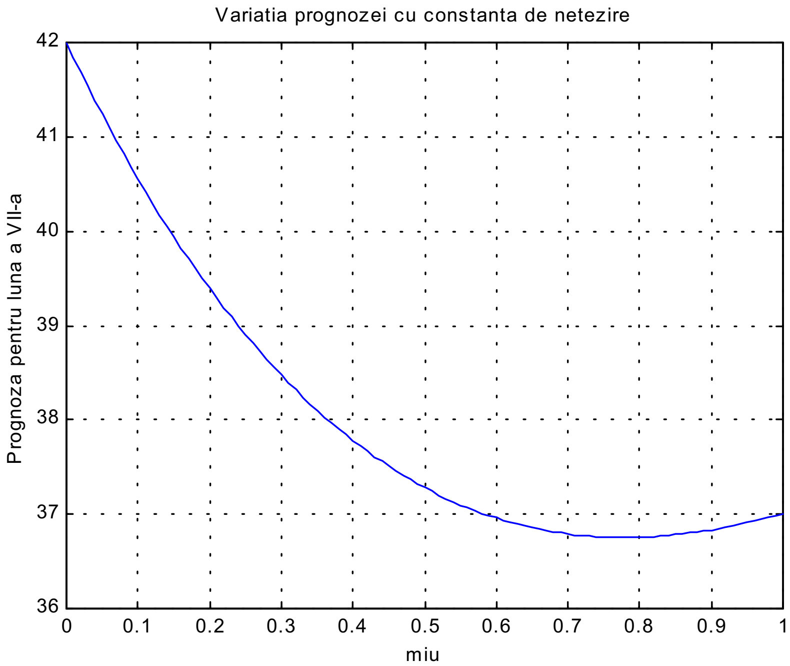
Pentru $\mu = 0{,}9$

$$\text{EMP} = [(42 - 41)^2 + (41{,}10 - 43)^2 + (42{,}81 - 38)^2 + (38{,}48 - 35)^2 + $$
$$+ (35{,}35 - 37)^2]/5 = 8{,}52$$

Cazul cu μ = 0,9 apare a da prognoze mai bune decât cel cu μ = 0,2 deoarece EMP este mai mică dacă μ = 0,9.
Pentru a reduce secventa de valori ale erorii la o valoare unică, cuprinzătoare s-a utilizat aici EMP. Mai sunt si alte modalităti de a judeca nivelul încrederii într-o prognoză. O altă valoare sintetică pe baza căreia se pot face judecvti de acest gen este eroarea medie absolută (EMA), care este suma erorilor luate în valoare absolută, raportată la numărul de erori.
Există metode care permit stabilirea valorii optime pentru constanta de netezire, adică a valorii μ care minimizează criteriul ales pentru aprecierea acuratetei prognozei, fie că este vorba de EMP, fie că este în discutie vreun alt criteriu. Pentru EMP, valoarea optimă în cazul datelor din această sectiune este μ = 0,86 la o valoare a EMP de 8,511. Căutarea acestei valori se poate face în moduri variate. O posibilitate este calculul direct, repetat.
Valorile optime ale constantei μ pot fi foarte diferite pentru criterii diferite. De pildă pentru EMA minimă se obtine μ = 0,59.



Revenind la criteriul erorii medii pătratice (EMP), este dat mai sus un grafic al variatiei EMP cu constanta de netezire μ. Graficul imediat următor aceluia evidentiază un fapt care nu poate fi trecut cu vederea: relativa stabilitate a valorii prognozate pentru o gamă de valori ale constantei de netezire μ destul de largă. Pentru $0,6 \le \mu \le 1,0$ prognoza se situează între 3675 si 3700 u.f. Curba este destul de plată în intervalul de valori μ mentionat. Faptul acesta arată că stabilirea unei valori potrivite pentru μ nu trebuie făcută foarte precis: a treia zecimală pare a fi aici de prisos.



Prognoze mai perfectionate prin serii temporale

Există metode mai perfectionate de elaborare a prognozelor pentru seriile temporale. Acestea sunt bazate pe modelele ARIMA (AutoRegressive Integrated Moving Average). În esentă aceste metode presupun că seriile temporale sunt generate de un proces probabilistic cu valori viitoare dependente de valorile trecute si de erorile de predictie trecute. Pentru a aplica modelele ARIMA seriilor temporale, acestea trebuie să fie stationare. O serie temporală este stationară dacă proprietătile statistice de genul mediilor, dispersiilor, autocorelatiilor sunt constante în timp. Dacă o serie temporală în forma initială nu este stationară, este posibil ca o functie de seria temporală să fie stationară: luând, de pildă, diferenta dintre valorile succesive din serie se poate obtine uneori o altă serie temporală care este stationară.

RETELE PETRI – MODELE PENTRU SISTEMELE DE PRODUCTIE FLEXIBILE

Introducere în teoria retelelor Petri

Retelele Petri sunt un instrument matematic de natură grafică datorat lui Carl Adam Petri. Aceste grafuri cu structură specială sunt utilizate la reprezentarea, modelarea si simularea unor sisteme foarte diverse în care dinamica evenimentelor, evolutiile paralele, dependentele conditionate (cum este sincronizarea), competitia pentru resurse etc. sunt nu numai prezente dar sunt si determinante. Fenomene de aceste tipuri apar frecvent în sistemele de productie, în protocoalele de comunicare, în calculatoare si în retelele de calculatoare, în programele în timp real, în sistemele de transport etc. Toate aceste sisteme sunt cunoscute în prezent ca *sisteme cu evenimente discrete*.

Un punct de vedere din unghiul teoriei grafurilor

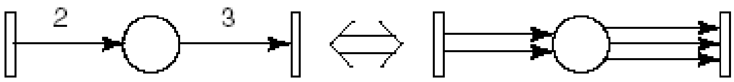
O retea Petri este o pereche ($\mathcal{G}$, M) compusă dintr-un graf bipartit orientat $\mathcal{G}$ = (E,V) si un marcaj initial M. Multimea nodurilor V este împărtită în două submultimi disjuncte, P si T. Elementele din P se numesc *pozitii*, elementele din T se numesc *tranzitii*. Pozitiile se notează P_i, i = 1,…, |P|, tranzitiile se notează T_j, j = 1,…, |T| (barele de modul exprimă numărul de elemente din multimea scrisă între ele sau, cum se mai spune, cardinalul acelei multimi). *Arcele* cuprinse în multimea E merg de la pozitii la tranzitii si de la trazitii la pozitii. Graful este bipartit si un arc nu poate uni o pozitie cu o pozitie si nici o tranzitie cu o tranzitie. În reprezentarea grafică pozitiile se reprezintă uzual prin cercuri, tranzitiile prin bare îngrosate (uneori prin dreptunghiuri). Arcelor li se atribuie ponderi, totdeauna numere întregi. Absenta grafică a ponderilor face subîntelese existenta unor ponderi unitare. Pentru o definire completă a unei retele Petri trebuie introdusă notiunea de marcaj initial. Marcajul initial atribuie fiecărei pozitii P_i un număr nenegativ M_i. La reprezentarea grafică acele numere sunt trecute, dacă e posibil, în cercurile care reprezintă pozitiile (sau stările). Vectorul coloană M cu componentele M_i se numeste marcajul initial al retelei. Se spune că pozitia P_i este *anterioară* tranzitiei T_j dacă există un arc de la P_i la T_j. Analog, se spune că pozitia P_i este *ulterioară* tranzitiei T_j dacă există un arc de la T_j la P_i.

Uzual, pozitiile reprezintă conditii, iar tranzitiile reprezintă evenimente. O tranzitie (un eveniment) implică un anumit număr de pozitii anterioare si

ulterioare, care reprezintă pre-conditii si post-conditii pentru acel eveniment. Dacă ponderile tuturor arcelor sunt egale cu unitatea, prezenta unui marcaj (denumit adesea si jeton) într-o pozitie se poate interpreta ca o conditie verificată asociată acelei pozitii. O altă interpretare mai generală este: M_i jetoane prezente în pozitia P_i indică o resursă disponibilă în cantitatea M_i.
Dintr-un punct de vedere clasic, marcajul unei retele Petri este identificat cu starea retelei. Schimbarea stării se produce după regulile care urmează:

- O tranzitie T_j poate fi abilitată si eventual amorsată, activată dacă orice pozitie anterioară acelei tranzitii contine atâtea jetoane cât este ponderea arcului care duce la tranzitia în discutie
- Când o tranzitie T_j este activată, din fiecare pozitie anterioară se consumă un număr de jetoane si, în consecintă, se diminuează numărul jetoanelor din acea pozitie exact cu numărul pondere a arcului care conectează pozitia la tranzitia respectivă; totodată, se adaugă pozitiilor ulterioare tranzitiei T_j atâtea jetoane câte sunt înscrise ca ponderi pe arcele emergente din T_j spre acele pozitii.

Observatie: în loc de a asocia ponderi arcelor, se poate face o reprezentare cu arce exclusiv cu pondere unitară; atunci între pozitii si tranzitii apar arce multiple în paralel.



Un punct de vedere din algebra liniară

Într-o analiză a marcajului si a pozitiilor, dacă se consideră vectorul (coloană) M al marcajului, se spune ca mai sus că M_i este numărul de jetoane din pozitia P_i. Fie o matrice de dimensiuni $|P| \times |T|$ notată D^- cu elementul generic d_{ij}^- egal cu ponderea arcului care pleacă din P_i si ajunge în T_j (arcele cu $d_{ij}^- = 0$ sunt inexistente). Analog, fie matricea D^+ de dimensiuni $|P| \times |T|$ cu d_{ij}^+ egal cu ponderea arcului de la tranzitia T_j la pozitia P_i (din nou, $d_{ij}^+ = 0$ semnalează arce care nu există). Pornind de la aceste definitii se spune că tranzitia T_j este abilitată si este amorsabilă dacă si numai dacă $M \geq D_{.j}^-$. Activarea tranzitiei produce un marcaj nou $\widetilde{M}$ care verifică ecuatia:

$$\widetilde{M} = M + D_{.j}^+ - D_{.j}^-$$

Dacă se defineste matricea $D = D^+ - D^-$ atunci se poate scrie

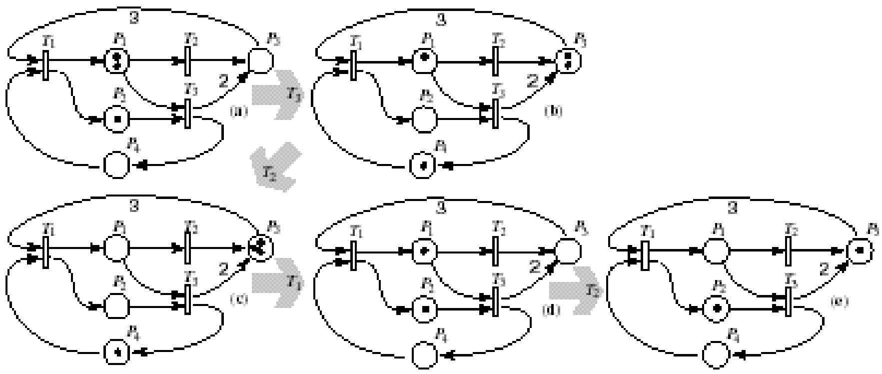
$$\widetilde{M} = M + Du$$

în care u este vectorul coloană definit ca $u_j = 1$, $u_i = 0$ pentru $i \neq j$.
Pentru mai multe tranzitii succesive, de pildă pentru două, se poate scrie

$$\tilde{\tilde{M}} = \tilde{M} + D\tilde{u} = M + Du + D\tilde{u} = M + D(u + \tilde{u}) = M + Du^+$$

cu u^+ un vector sumă a vectorilor u asociati unor tranzitii simple, în particular două, un vector care nu poate avea componente negative.

Observatie: Existenta unui vector de componente nenegative u astfel ca $\tilde{M} = M + Du$ nu implică obligatoriu posibilitatea de a obtine marcajul $\tilde{M}$ din marcajul M, prin una sau mai multe tranzitii. Conditia $M' \geq D^-_{.j}$ trebuie să se verifice la fiecare pas intermediar când un marcaj M' trece la marcajul M'' prin executarea unei anumite tranzitii T_j. În plus, în cazul succesiuni de tranzitii, vectorul u nu spune nimic relativ la ordinea în care tranzitiile trebuie să aibă (au) loc, ceea ce este foarte important. Datorită acestor restrictii sistemul nu este realmente liniar si principiul suprapunerii efectelor nu se verifică decât ocazional.



Un exemplu: Pentru reteaua Petri din figura alăturată, conform definitiilor enuntate

$$D^- = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 3 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},\ D^+ = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \text{ si } D = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ -3 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Săgetile (cenusii) din figură indică stările succesive ale retelei după executarea tranzitiilor înscrise pe acele săgeti. După executarea secventei de tranzitii $T_3T_2T_1T_2$ toate executabile în ordinea mentionată, se obtine marcajul

$$\tilde{M} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + D\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + D\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Semantica retelelor Petri

Componentele diverse ale unei retele Petri au semnificatiile care urmează:

- Marcajele reprezintă resurse în deplinul sens al cuvâtului. Poate fi vorba de resurse fizice, cum sunt cele materiale, sau de informatii, mesaje, conditii etc.
- Din pozitii se asteaptă accesul la anumite resurse
- Tranzitiile reprezintă actiuni consumatoare de resurse spre a fi prelucrate, tranzitiile sunt producătoare de alte resurse
- Ponderile arcelor care leagă o pozitie cu o tranzitie reprezintă numărul minim de resurse din categoria stocată în acea pozitie necesar pentru a executa tranzitia
- Ponderile arcelor care unesc o tranzitie cu o pozitie reprezintă numărul exact de resurse noi din categoria celor stocate în acea pozitie, resurse produse prin actiunea definită de tranzitie
- Numărul total de marcaje, de jetoane dintr-o retea Petri nu se conservă obligatoriu: se pot imagine actiuni de asamblare, se pot imagina actiuni de demontare a unor ansamble în părti componente; mesajele pot fi combiante pentru obtinerea unui mesaj nou (cum este cazul însumării a două numere) sau un mesaj dat poate fi difuzat către mai multe pozitii

Invarianti

Invarianti în pozitii

Se admite că v este un vector linie de dimensiune $|P|$ care verifică relatia $vD = 0$. Atunci, produsul vM, produs care se poate interpreta ca o sumă ponderată a valorilor marcajului cu ponderi egale cu componentele vectorului v se mentine constant oricare ar fi secventa de tranzitii executată. De pildă pentru un marcaj $\widetilde{M}$ rezultat din M are loc implicatia evidentă

$$\widetilde{M} = M + Du \Rightarrow v\widetilde{M} = vM + vDu = vM$$

Exemplu. Pentru reteaua Petri dată mai devreme se observă că

$$[0 \quad 1 \quad 0 \quad 1]D = [0 \quad 1 \quad 0 \quad 1]\begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ -3 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix} = [0 \quad 0 \quad 0 \quad 0]$$

Asta se traduce prin aceea că numărul total al marcajelor din pozitiile P_2 si P_4 se mentine constant independent de tranzitiile executate.

Invarianti pentru tranzitii

Fie acum un vector coloană u de dimensiune $|T|$ si cu componente nenegative care verifică egalitatea $Du = 0$. Atunci, oricare secventă de tranzitii fezabilă reprezentată de vectorul u conservă marcajul initial. Într-adevăr

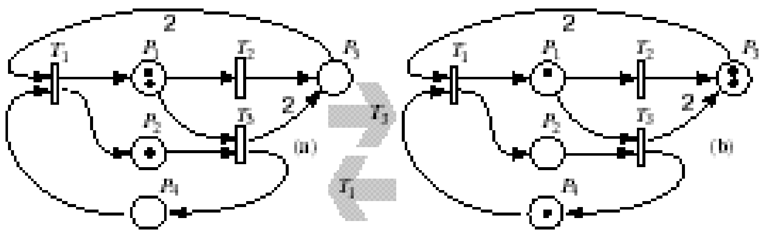
$$\widetilde{M} = M + Du = M$$

Cum s-a discutat mai sus, secventa de tranzitii fezabile cuprinsă în u poate si să nu existe.

Exemplu. Pentru reteaua Petri din figura alăturată

$$Du = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ -2 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Se poate verifica faptul că vectorul $u = [1 \quad 0 \quad 1]^{T}$ ar putea reprezenta fie secventa T_3T_1, fie secventa T_1T_3 dar numai una din ele este fezabilă.



Conflicte

Definitie. Se spune că două tranzitii diferite T_i si T_j sunt în *conflict structural* dacă

$$\exists k : D^{-}_{ki} \times D^{-}_{kj} \neq 0$$

ceea ce înseamnă că pozitia P_k premerge ambele tranzitii. Se spune despre un conflict structural că este si *efectiv* dacă, în plus, la marcajul M ambele tranzitii pot fi activate, adică atât $M \geq D^{-}_{.i}$ cât si $M \geq D^{-}_{.j}$.

Exemplu. În reteaua Petri de mai sus tranzitiile T_2 si T_3 sunt în conflict structural (pozitia P_1 este anterioară ambelor tranzitii). Acest conflict este si efectiv în prima parte a figurii (a), dar nu este efectiv după ce tranzitia T_3 s-a produs, în partea din dreapta a figurii (b).

Termenul de conflict provine din aceea că dacă tranzitiile T_i si T_j sunt în conflict structural efectiv atunci nu se poate produce, nu se poate amorsa decât una din aceste tranzitii si nu ambele deoarece nu sunt suficiente jetoane în pozitiile anterioare oricum s-ar succeda tranzitiile în discutie. În situatiile de acest gen este necesară o decizie: care din cele două tranzitii se amorsează. Altfel spus, două tranzitii în conflict strucutural sunt în competitie pentru resusele accesibile cel putin într-o pozitie anterioară pe care o împart.

Paralelism

Definitie. Se spune că două tranzitii T_i si T_j sunt în *structural paralele* dacă

$$D_{.i}^{-T} \times D_{.j}^{-} = 0$$

ceea ce înseamnă că tranzitiile T_i si T_j nu au pozitii anterioare în comun. Se spune că paralelismul structural este *efectiv* la un marcaj M dacă, în plus, ambele tranzitii pot fi activate (fiind deja abilitate), adică atât $M \geq D_{.i}^{-}$ cât si $M \geq D_{.j}^{-}$.

Exemplu. În reteaua Petri de mai devreme tranzitiile T_1 si T_2 sunt structural paralele. Acest paralelism este efectiv în cazul din figura secundă, cea din dreapta (b), ceea ce nu este cazul cu prima figură (a).

Viabilitate

Definitie. Tranzitia T_i este viabilă (sau vie) atunci când oricare ar fi marcajul accesibil din marcajul initial, există o secventă de tranzitii fezabile, care conduce la un marcaj pentru care tranzitia este abilitată pentru amorsare. O retea Petri se spune că este viabilă dacă toate tranzitiile ei sunt viabile.

De notat că dacă o tranzitie este viabilă atunci ea poate fi amorsată indefinit (se zice că există o suceesiune de tranzitii fezabile prin care acea tranzitie este vie la nesfârsit). Dacă o tranzitie nu este viabilă atunci există posibilitatea ca reteaua Petri să functioneze numai un timp finit dar prin repetarea unei grupe de tranzitii poate fi în functiune si un timp indefinit.

Exemplu. Reteaua Petri din figura cu cinci faze dată mai sus nu este vie deoarece secventa de tranzitii considerată conduce la un marcaj pentru care nici o tranzitie nu mai este abilitată pentru executie (situatie de blocaj, “dead-lock”).

Mărginire, sigurantă

Definitie. O pozitie P_i este k-mărginită dacă marcajul ei nu este (nu poate fi) mai mare decât k, oricare ar fi marcajul (accesibil). O pozitie se numeste sigură dacă este 1-mărginită. O retea Petri este sigură dacă toate pozitiile sale sunt 1-mărginite (sigure).

Notăm că dacă o pozitie este de tipul magazie, buffer cu o capacitate finită k (un depozit, de pildă) atunci în mod necesar reteaua Petri trebuie să fie k-mărginită dacă modelarea este corect făcută. Mai departe se va da o metodă simplă de asigurare a mărginirii corecte.

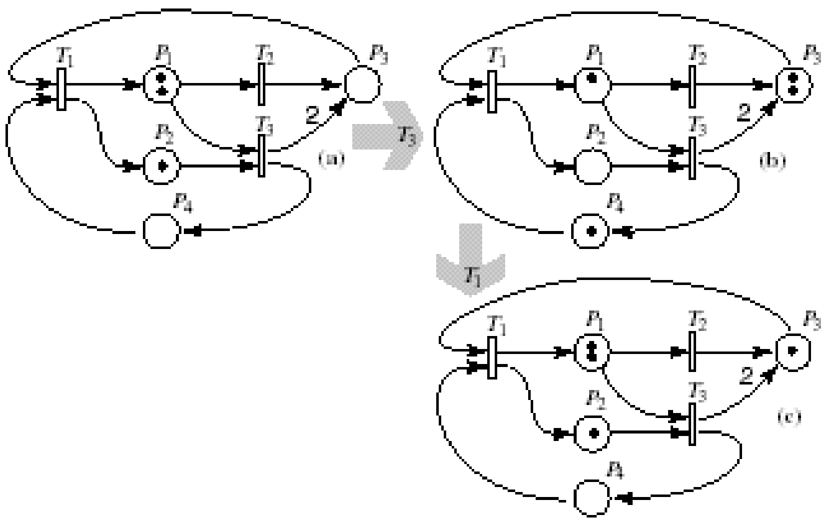
Exemplu. Secventa de tranzitii T_3, T_1 din figura care urmează conduce la un marcaj care coincide cu cel initial cu exceptia celui pentru pozitia P_3 care are un jeton în plus. Mai mult, dacă secventa de tranzitii mentionată se repetă de k

ori în pozitia P_3 se acumulează k jetoane si, în consecintă, pozitia nu este mărginită.
Sub aspect matematic se scrie (cu u vectorul asociat secventei T_3, T_1):

$$Du = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ -1 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

si atunci

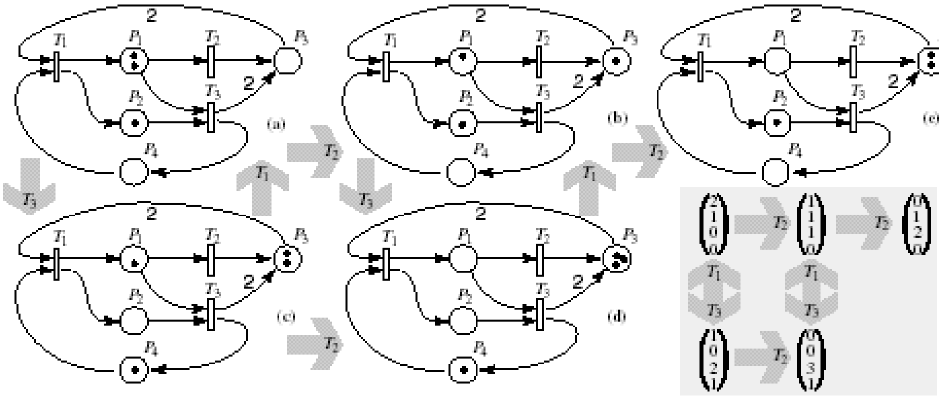
$$\widetilde{M} = M + Du > M$$



Prin repetarea secventei de tranzitii mentionate, permanent fezabilă, se obtine o crestere a marcajului pozitiei P_3 indefinită. Pornind de la acest exemplu se poate observa că ori de câte ori o retea Petri admite o secventă de tranzitii fezabilă pentru care vectorul u verifică relatia $Du > 0$, reteaua Petri este nemărginită (rationamentul pentru cazul general este similar celui din exemplul de mai sus).

Marcaje accesibile

Majoritatea proprietătilor de mai sus se pot verifica dacă se cunoaste multimea marcajelor accesibile pornind de la marcajul initial. Desigur, calculul tuturor marcajelor accesibile din marcajul initial nu este în general o sarcină usoară. Figura care urmează ilustrează un asemenea calcul într-un caz simplu.



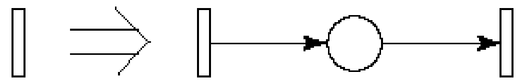
Timpi asociati cu pozitiile si tranzitiile

Teoria originală a retelelor Petri tratează ordinea evenimentelor si, în consecintă, diamica retelei a fost considerată ca o succesiune de evenimente (de tranzitii) restrânsă numai la considerente de logică (o tranzitie se poate desfăsura numai dacă este abilitată). În acest context întrebarea "cât timp consumă un eveniment?" nu se pune. Dar pentru a răspunde la întrebări relativ la performantele retelei si sistemului modelat cu reteaua (de pildă, "cât de repede poate produce sistemul?") este necesar a introduce în discutie timpul. Se pot într-adevăr asocia cu pozitiile si cu tranzitiile durate pe calea următoare:

- Durate asociate cu pozitiile: duratele minime pentru ca jetoanele să devină permanente într-o pozitie, înainte de fi capabile a contribui la abilitarea (si amorsarea) unei tranzitii următoare. Duratele acestea se numesc *timpi de asteptare*.
- Durate asociate tranzitiilor: durate care separă începutul (consumul de jetoane din pozitiile premergătoare) si finalizarea actiunii (producerea de jetoane destinate pozitiilor următoare) corespunzătoare tranzitiilor. Aceste durate au primit numele de *timpi de actiune*.

Duratele de executie pot fi utilizate, de pildă, pentru a reprezenta timpul de productie în cazul sistemelor de productie (unde tranzitia reprezintă timpul uzual consumat pe masina unealtă). Timpii de asteptare ar putea reprezenta timpii de transport (în cazul în care o pozitie reprezintă o rută sau un canal prin care comunică două procese) sau la fel de bine timpul minim de acces, necesar accesibilitătii (cum ar fi timpul de răcire al unei piese trecute printr-un cuptor, înainte de a i se putea aplica prelucrarea următoare). Timpii, duratele de asteptare si de executie pot fi constante sau variabile, pot fi deterministe sau aleatoare. Nu trebuie ignorat nici timpul de constituire a numărului de jetoane dintr-o pozitie necesar abilitării unei tranzitii.

Observatie. În realitate, fără pierdere din generalitate, se poate admite că toate actiunile sunt instantanee (toate tranzitiile se petrec în timp nul). Tranzitiile cu durată nenulă se divid în două tranzitii instantanee (începutul si terminarea actiunii) separate de o pozitie care are timpul de asteptare egal cu timpul de executie al tranzitiei originare (v.figura alăturată).



Tranzitii de intrare si de iesire

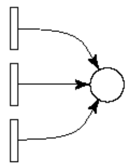
Tranzitiile care nu au pozitii premergătoare se numesc *tranzitii de intrare* sau surse. Actiunile de acest tip depind de decizii externe, sunt controlate din exterior. Tranzitiile care nu au pozitii următoare se numesc tranzitii de iesire sau consumatori. Tranzitiile de acest gen indică producerea de jetoane destinate exteriorului.
Aceleasi definiti se pot aplica si pozitiilor: pozititile de intrare trebuie aprovizionate cu jetoane din exterior. În realitate, cum se va vedea mai departe, în acord cu clasa particulară a retelei Petri în studiu, poate rezulta mai potrivită utilizarea la periferie a tranzitiilor (si nu a pozitiilor de intrare si de iesire) sau a pozitiilor (si nu a tranzitiilor de intrare si de iesire).

Reguli de functionare

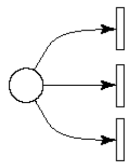
Până aici s-au impus executării tranzitiilor numai restrictii de ordin logic fără a specifica momentul în care o tranzitie este si executată. Acum, că s-a adus în discutie timpul, se poate defini regula de functionare cunoscută ca regula *timpului de actiune cel mai apropiat*: tranzitiile se execută cât de prompt posibil, adică deîndată ce sunt asigurate toate jetoanele necesare pentru a abilita trazitia.
În imediată legătură cu regula de mai sus se introduc *reguli de prioritate*, regulile de arbitraj în cazul pozitiilor implicate într-un conflict sau modalitatea de a indica ce tranzitie trebuie să se execute atunci când apare un conflict si *traiectoriile la intrare*, functii $u_i : N \rightarrow R^+$ (una pentru fiecare tranzitie de intrare U_i) cu $u_i(n)$ instanta în care tranzitia de intrare U_i se află la momentul n.
Cu aceste reguli de functionare se pot determina toate momentele la care se produc evenimentele din retea ca, de exemplu, actiunile succesive, sosirea si plecarea unor jetoane într-o/dintr-o pozitie etc. În particular, se ajunge la momentele când se execută tranzitiile de iesire ceea ce constituie *traiectoriile la iesire*.

Competitie si sincronizare

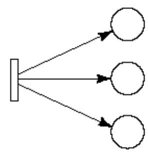
Competitia pentru a produce, reunirea într-o pozitie. Această situatie se întâlneste atunci când o pozitie are mai multe tranzitii premergătoare. În acest caz există mai multe surse care produc jetoane destinate acelei pozitii (v.figura). Ca exemplu de acest tip poate servi cazul unei pozitii-depozit unde sosesc produse ale mai multor masini, tranzitiile reprezentând tocmai actiunile acestor masini.



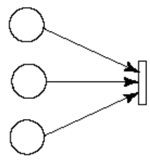
Competitia pentru a consuma, ramificarea dintr-o pozitie. În acest caz o pozitie are mai multe tranzitii care o urmează (v.figura următoare). Tranzitiile cunt în competitie pentru jetoanele acestei pozitii. Situatia se tratează ca un conflict structural cum s-a discutat mai devreme.



Sincronizarea în a produce, ramificarea dintr-o tranzitie. Aici o tranzitie are mai multe pozitii următoare (v.figura). În aceste cazuri jetoanele (resurse, părti componente, mesaje etc.) sunt emise simultan către pozitiile consumatoare următoare.



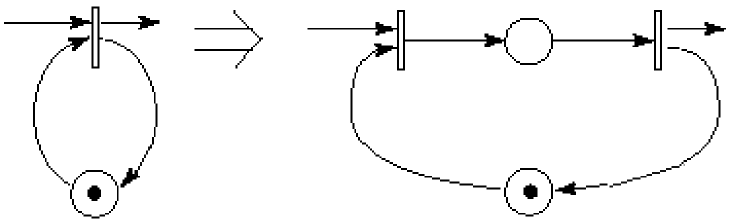
Tranzitia ar putea corespunde, de pildă, unei operatii de dezmembrare a unei pese în părtile ei componente.
Sincronizarea în consum, reunirea într-o tranzitie. Situatia corespunde cazului în care o tranzitie are mai multe pozitii premergătoare.



Jetoanele asteaptă în acele pozitii momentul în care apare ultimul jeton care abilitează tranzitia (se spune că starea fiecărei pozitii durează cel putin cât durata de asteptare a celei pozitii). În acel moment se consumă concomitent toate jetoanele necesare pentru activarea tranzitiei.

Mecanisme de control

Prevenirea activării multiple simultane a unei tranzitii. În conformitate cu definitiile de mai devreme, nimic nu împiedică o tranzitie să devină activă simultan de mai multe ori: dacă activarea unei tranzitii nu este instantanee atunci se poate întâmpla ca tranzitia să se amorseze da mai multe ori înainte ca actiunea determinată de prima activare să se fi isprăvit. Asta ar însemna ca o masină destinată efectuării unei anumite operatii pe un tip de piese, pe rând pentru fiecare piesă, să fie "inundată" de alte piese similare care, evident, nu pot fi servite paralel. Pentru a preveni o situatie de acest gen se atasează acelei tranzitii o pozitie suplimentară. Această pozitie trebuie să aibă ca unică tranzitie anterioară si unică tranzitie următoare tranzitia în discutie. Figura care urmează explică metoda în două variante echivalente. În varianta din stânga tranzitia are durata *t*, durată necesară executării actiunii căreia îi corespunde. Pozitia suplimentară are durata θ de punere în miscare a actiunii. În varianta din dreapta tranzitia a fost descompusă în două tranzitii instantanee cu o pozitie între ele cu durata de asteptare *t*. Pozitia suplimentară are aceleasi caracterisitici.



Agregatul mai simplu sau mai complex al buclei create se numeste *reciclarea* tranzitiei. În plus, se poate atribui un timp de asteptare pozitiv pozitiei reciclante pentru a forta un timp minim între finalizarea unei actiuni si initierea următoarei (se poate vorbi de un timp de punere în miscare/în functiune). Se observă că pozitia jetonului în buclă indică dacă tranzitia este ocupată sau liberă, ocuptă atunci când jetonul din pozitia suplimentară lipseste.

Controlul fluxului. O modificare similară permite limitarea fluxului de jetoane printr-o tranzitie cu timp de actiune nul. Se observă (v.figura următoare) că dacă marcajul initial al pozitiei suplimentare asociate tranzitiei (pozitie care, de asemenea, trebuie să aibă ca unică tranzitie premergătoare si următoare tranzitia considerată) este *m* si timpul ei de asteptare *t* atunci fluxul maxim de jetoane prin acea tranzitie este de *m* jetoane la fiecare *t* unităti de timp.

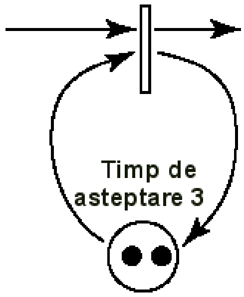
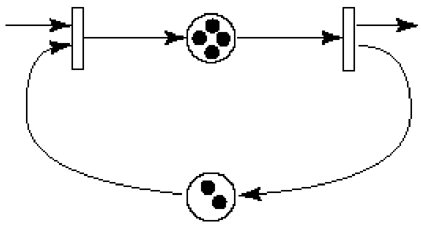


Figura indică un debit maxim de două jetoane la fiecare trei unităti de timp.

Pozitii cu capacitate limitată. Modelarea unor sisteme fizice pune problema practică a capacitătii limitate a unor pozitii. Există firesc o limită superioară a numărului de jetoane pe care o pozitie le poate contine. Mărginirea specificată si sigură a unei pozitii se poate obtine pe baza următorului algoritm:

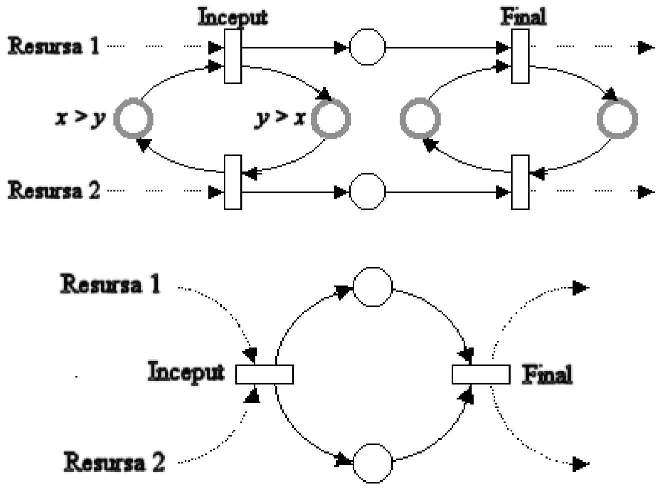
1. Pentru o pozitie p k-mărginită se adaugă o pozitie suplimentară p' cu marcajul initial $M(p') = k - M(p)$
2. Între fiecare tranzitie t si pozitiile suplimentare de genul p' se definesc arce suplimentare cu ponderile $w(t, p') = w(p, t)$ si $w(p', t) = w(t, p)$ ceea ce face ca suma jetoanelor din pozitia p si din pozitia complementara p' să fie aceeasi si înainte su după executarea unei tranzitii.

Figura alăturată este un exemplu.



Este aici vorba de un depozit intermediar între două servicii marcate prin tranzitiile din figură. Capacitatea depozitului este de maximum 6 unităti.

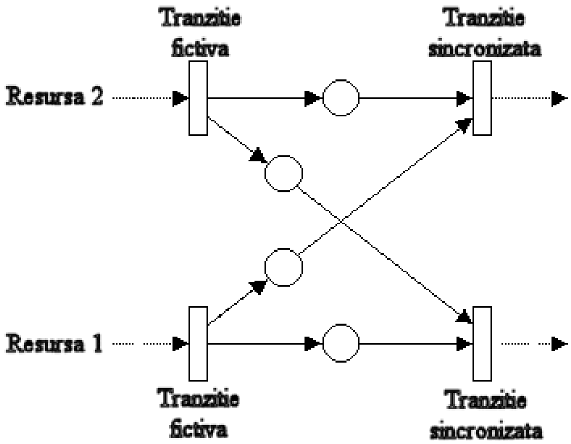
Sincronizarea activării tranzitiilor. Uneori se poate întâmpla ca două sau mai multe tranzitii să reprezinte aceeasi actiune fizică. Într-un asemenea caz tranzitiile trebuie să se *sincronizeze* pentru a se amorsa simultan. Asta se poate realiza cel putin în două moduri care duc la un gen de "unire" a tranzitiilor considerate (Unul din cele două moduri nu este deplin acceptabil sub incidenta teoriei clasice a retelelor Petri; cum se va arăta mai departe, sub aspect matematic modul acela este totusi corect si adecvat în a exprima simultaneitatea). Este vorba de a face să coincidă începutul si sfârsitul unei etape pentru mai multe resurse implicate simultan într-o anumită etapă. Se apelează la "circuite de sincronizare" fără temporizare si fără jetoane. Fiecare din cele două arce ale circuitului de sincronizare include si impune câte o inegalitate, una de sens opus celeilalte, la momentele de activare a tranzitiilor, de unde egalizarea momentelor de activare ale tranzitiilor. Aceste tranzitii pot apoi să fie puse laolaltă, pot fuziona (v.figurile următoare).



Existenta de circuite fără jetoane (si fără temporizare), acceptabilă sub aspect matematic, este contrară regulilor ortodoxe ale retelelor Petri. Se poate justifica functionarea spunând că se "împrumută" jetoanele (absente din circuitul de sinronizare) pentru a activa tranzitiile si că schema este în măsură a restitui aceste jetoane într-un timp nul. Fuziunea tranzitiilor sincronizate înlătură orice discutie.

De notat că un număr egal de săgeti intră în si ies din tranzitiile sincronizate. În consecintă, numărul total de jetoane din graf (si nu numai din circuite) se conservă în timpul functionării. Se recuperează de asemenea interpretarea de "resurse" a jetoanelor însesi.

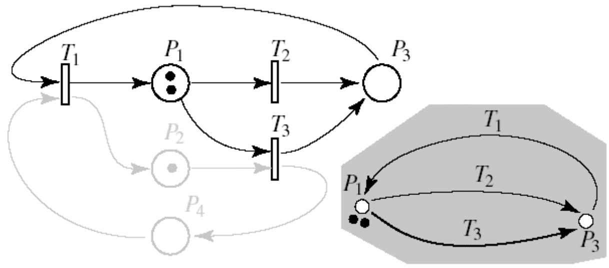
O altă solutie foarte diferită permite si aceasta sincronizarea a două tranzitii. Această solutie evită circuitele de sincronizare cu pretul introducerii unor tranzitii fictive înaintea tranzitiilor adevărate. Solutia e ilustrată în figura alăturată. Se poate verifica prin simularea functionării retelei Petri si, mai departe, prin ecuatii, că sincronizarea este efectivă.



Această diversitate de solutii grafice produs al aceleiasi ecuatii matematice este o ilustrare a interesului de a a pune în ecuatii grafurile de evenimente.

Retele Petri speciale. Asincronism, masini de stare si automate

Retelele Petri asincrone sunt acelea în care toate tranzitiile au cel mult o pozitie anterioară si cel mult o pozitie următoare (v.figura care urmează). În asemenea retele nu există tranzitii de intrare si de iesire si, de aceea, "terminalele" sunt de tipul pozitiilor (ceea ce face ca fiecare tranzitie să posede exact o singură pozitie premergătoare si o singură pozitie următoare).
Retelele Petri cu toate tranzitiile având exact o pozitie premergătoare si exact o pozitie următoare se numeste *masină de stare*. În masinile de stare asignarea timpului pentru tranzitii si pozitii nu este importantă. Singurul efect al atribuirii este de întârziere a executării tranzitiilor. Aici problema principală este cea logică (accesibilitatea marcajelor, eliminarea blocajelor etc.). În general efortul principal de control este orientat pe executarea tranzitiilor. Când numărul total de marcaje este unu, gândul poate duce la faptul că acel marcaj unic arată starea sistemului (pozitiile reprezintă stările posibile ale sistemului) si reteaua obtinută se poate interpreta ca fiind un automat. Dacă în plus fiecare pozitie are exact o tranzitie următoare acel automat rezultă a fi determinist. Dacă nu acesta este cazul automatul nu este determinist (v.figura) si atunci pentru fiecare stare sunt posibile traiectorii diferite. În cazul *non*-determinist se atribuie probabilităti arcelor care pleacă dintr-o pozitie atunci se obtine un automat stochastic. Partea pe fond cenusiu din figura de mai jos detaliază căile alternative de a ajunge de la pozitia P_1 la pozitia P_3.



Invarianti. Cum într-o masină de stare fiecare tranzitie are o pozitie premergătoare si una următoare, matricea $D = D^+ - D^-$ contine pe coloana asociată cu tranzitia un -1 si un 1 (dacă arcele toate au ponderea unitară). În realitate matricea D poate fi considerată o matrice de incidentă noduri-arcuri în graful orientat, care se obtine dacă fiecare tranzitie se înlocuieste cu un arc care leagă pozitia anterioară de pozitia următoare acelei tranzitii (nodurile acestui graf sunt pozitiile grafului initial). Cu această observatie si cu rezultatele simple din teoria grafurilor se obtin consecintele care urmează.

- *Invarianti pentru pozitii*: deoarece matricea D are structura coloanelor arătată (un -1 si un $+1$) rezultă că

$$(\ldots \quad 1 \quad \ldots \quad 1 \quad \ldots)D = 0$$

motiv pentru care numărul total de jetoane într-o masină de stare este permanent acelasi.

Pentru ca o retea Petri să fie viabilă este necesar ca marcajul initial să nu fie nul. Pentru o masină de stare această conditie este si suficientă dacă structura este conexă.

- *Invarianti pentru tranzitii*: dacă u este un vector coloană caracteristic al unui circuit (matricea de incidentă arcuri-noduri este transpusa matricei de incidentă noduri-arce) se spune că componentele lui u care corespund arcelor (tranzitiilor) unui circuit au valoarea 1 si componentele celelalate sunt nule. Atunci se verifică relatia

$$Du = 0$$

Grafuri cu evenimente temporizate

Se consideră grafurile de evenimente temporizate (GET) cu ponderile arcelor unitare si temporizare constantă si numai pentru pozitii. Se demonstrează că sistemele de acest gen pot fi modelate ca sisteme "liniare" într-o semnificatie diferită a termenului.

Punctul de vedere "dater". Este convenabil a se considera că GET sunt delimitate de tranzitii adică toate pozitiile au tranzitii premergătoare si următoare. Această conventie nu implică vreo restricitie deoarece:

- Orice pozitie de intrare îsi ia jetoanele din exterior si despre ele se poate gândi ca provenind de la o tranzitie premergătoare controlată dinafară.

- Orice pozitie de iesire poate fi urmată de o tranzitie care se activează numai pentru jetoanele care sosesc la acea pozitie.

Se notează tranzitiile de intrare activate cu u_j cu $j = 1, \ldots, m$, etc. cu indicele j asociat unei ordonări temporale. În formă similară se notează tranzitiile de iesire activate cu y_l, $l = 1, \ldots, p$ si tranzitiile interne activate cu x_i, $i = 1, \ldots, n$. Din cauza succesiunii lor în timp aceste numere functii de timp se numesc dater-e (cele care datează, care fac calendarul).

Cu adoptarea regulei că tranzitiile sunt amorsate imediat ce este posibil si cu mentiunea că orice conflict este absent, singurele elemente necesar a fi cunoscute pentru a realiza o simulare sunt:

- Momentele când sunt activate tranzitiile de intrare (prin decizii externe) pe întreaga durată a simulării
- Momentele când sunt disponibile jetoanele prezente în marcajul initial (se poate considera că aceste jetoane sunt prezente de un anumit timp, înainte de a începe simularea)

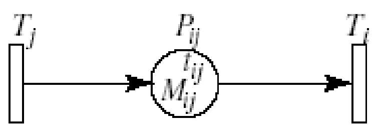
Cunoscând aceste informatii, este posibil a determina când se vor produce tranzitiile interne si tranzitiile de iesire.

Ecuatiile dater. Se prezintă dater-ele asociate cu fiecare tranzitie. Pentru o tranzitie x_i, variabila asociată $x_i(k)$ se interpretează ca momentul în care se produce cea de a k amorsare. De la începutul simulării activările succesive ale unei tranzitii sunt numărate secvential de la o origine generală (uzual zero, dar poate fi si un număr negativ). Asadar, functia $k \to x_i(k)$ este nedescrescătoare (unele activări se pot produce simultan si, de aceea, functia poate să nu fie strict crescătoare).

Timpul se poate observa pe o scară *reală*, *ratională* sau *întreagă*, de la caz la caz, $x(k) \in R$, Q sau Z.

Ecuatiile de datare (daters) rezultă din consideratiile următoare:

- Dacă tranzitia x_i este ulterioară tranzitiei x_j si este separată de o pozitie notată P_{ij}, atunci cea de a k executare a tranzitiei x_i va consuma jetonul produs de executarea numărul $k - M_{ij}$ a tranzitiei x_j cu M_{ij} marcajul initial al pozitiei P_{ij}.
- Dacă timpul de asteptare în pozitia P_{ij} este t_{ij}, executarea numărul k a tranzitiei x_i nu se poate produce dacă nu s-a scurs cel putin t_{ij} unităti de timp de la amorsarea numărul $k - M_{ij}$ a tranzitiei x_j.



- Tinând cont de această relatie pentru toate tranzitiile x_j anterioare tranzitiei x_i, maximul tuturor acestor momente determină momentul amorsării numărul k a tranzitiei x_i.

În final, ntând cu $*i$ multimea de indici ai tranzitiilor anterioare tranzitiei x_i, ecuatia fundamentală pentru GET este

$$x_i(k) = \max_{j \in *i} \left(x_j(k - M_{ij}) + t_{ij} \right)$$

Ecuatiile sunt valide totdeauna, chiar si când jetoanele considerate au fost produse prin activarea tranzitiilor în timpul simulării. Dacă numărătoarea activărilor încep cu $k = 0$ ecautiile se validează pentru $k \geq M_{ij}$. Ecuatiile sunt valide fără restrictii când jetoanele marcajului initial nu contribuie la operatia de luare a maximumului.

Jetoanele marcajului initial este *lista de utilizare* la momentul acoperitor $-\infty$. Se vorbeste atunci de *conditii initiale canonice*.

În continuare se expune cum actionează conditiile initiale arbitrare (nu neapărat canonice).

Din ecuatia generică de mai sus, validă cu restrictia din paragraful anterior, rezultă în mod evident că forma generală a ecuatiilor de datare pentru un GET complet este următoarea (pentru explicatii privind operatorii din relatiile prezentate, a se citi NOTA de la sfârsitul acestei sectiuni)*:

$$x(k) = A_0 x(k) \oplus A_1 x(k-1) \oplus \ldots \oplus B_0 u(k) \oplus B_1 u(k-1) \oplus \ldots$$

$$y(k) = C_0 x(k) \oplus C_1 x(k-1) \oplus \ldots \oplus D_0 u(k) \oplus D_1 u(k-1) \oplus \ldots$$

în care

- $x(.)$, $u(.)$ si $y(.)$ sunt vectori coloană de dimensiuni n, m, p
- A_i, B_i, C_i, D_i, sunt matrici de dimensiunile $n \times n$, $n \times m$, $p \times n$ si $p \times m$. Numărul maxim de matrici (nenule) din fiecare tip este egal cu maximumul marcajului initial al pozitiilor din GET, cum se explică în continuare
- Regula căreia i se supune elementul (r, s) al matricei A_i este: dacă r este o tranzitie internă imediat ulterioară tranzitiei interne s si dacă are i jetoane în marcajul initial al pozitiei P_{rs} atunci elementul $(A_i)_{rs}$ nu este nul (adică este distinct de ε) si este egal cu *timpul de asteptare* al pozitiei P_{rs}. Cu alte cuvinte dacă se consideră graful GET cu tranzitiile ca noduri si cu pozitiile ca arcuri si se mentin numai nodurile interne si arcele cu exact i jetoane initiale atunci acesta este *graful de precedentă* al tranzitiilor, cu ponderi pe arce egale cu timpii de asteptare al pozitiilor corespunzătoare.
- De o forma asemănătoare, B_i se bazează pe un graf care mentine numai nodurile corespunzătoare tranzitiilor de intrare si interne si arcele cu exact i jetoane initiale dintre o tranzitie *de intrare* si o tranzitie *internă*; de data aceasta este vorba de *graful de tranzitie* corespunzător.
- De formă analogă, C_i se bazează pe un graf care mentine nodurile interne si de iesire si arcele cu exact i jetoane initiale dintre o tranzitie *internă* si o tranzitie *de iesire* fiind acesta graful de graful de tranzitie al acestei matrici.
- Matricea D_i se defineste la fel cu cele precedente; se mentin numai graful cu nodurile de intrare si de iesire cu arce cu exact i jetoane initiale
- Algebra utilizată este algebra max-plus

- *Conditiile initiale* sunt $x(k) = \varepsilon$ pentru orice k negativ ceea ce reflectă supozitia că prima activare a fiecărei tranzitii care modifică marcajul initial al pozitiei anterioare este zero.

O formă canonică. Ecuatiile de mai sus sunt implicite deoarece variabilele $x(k)$ sunt prezente în ambii termeni ai relatiei prime. Aceste ecuatii se pot rezolva. Fără a intra în detalii, rezultatul este

$$x(k) = A_0^*\left(A_1 x(k-1) \oplus \ldots \oplus B_0 u(k) \oplus B_1 u(k-1) \oplus \ldots\right)$$

Această formă permite o examinare mai atentă din punct de vedere practic. S-a luat mai sus solutia minorantă. Întrebare: dacă aceasta nu-i unică, ce efect are această alegere? Răspunsul este în relatie cu cele două reguli ale jocului:

1. Tranzitiile se activează *deîndată ce este posibil*, ceea ce face compatibil dater-ul cel mai mic posibil cu ecuatiile
2. Ecuatiile implicite sunt valide în virtutea influentei marcajului initial; se selectionează conditiile initiale de asa natură încât oricare altă alegere poate numai să întârzie evenimentele ulterioare.

Relatiile de mai sus sunt foarte asemănătoare cu ecuatiile care descriu un sistem în varianta ecuatie-de-stare – ecuatie-de-observare si multe rezultate din teoria sistemelor se pot aplica aici schimbând doar regulile de calcul conform algebrei dioidului R_{max}.

---

[*] NOTA: Relatiile sunt scrise într-o algebră specială, *algebra dioizilor*.

Pe multimea numerelor reale se defineste o structură algebrică de *dioid*, descrisă pe scurt imediat.

Definitie: Un dioid este o multime **D** dotată cu două operatii notate “$\oplus$” si “$\otimes$”, care se numesc respectiv “adunare” si “multiplicare” si care verifică axiomele:

Adunarea este asociativă: $(a \oplus b) \oplus c = a \oplus (b \oplus c)$

Adunarea este comutativă: $a \oplus b = b \oplus a$

Adunarea admite un element neutru (notat ε si denumit “zero”): $a \oplus \varepsilon = a$

Multiplicarea este asociativă: $(a \otimes b) \otimes c = a \otimes (b \otimes c)$

Multiplicarea admite un element neutru (notat e si denumit “identitate”): $e \otimes a = a \otimes e = a$

Multiplicarea este distributivă fată de adunare: $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$ si analog pentru multiplicarea sumei la dreapta

"Zero"-ul este absorbant pentru multiplicare: $\varepsilon \otimes a = a \otimes \varepsilon = \varepsilon$

Adunarea este idempotentă: $a \oplus a = a$

Ca si în algebra uzuală, semnul de multiplicare este uneori omis. Dioidul este comutativ dacă operatia de multiplicare este si ea comutativă. Un *sub-dioid* este o submultime de elemente ale unui dioid stabilă la operatiile "$\oplus$" si "$\otimes$" si care contine elementele speciale ε si e.

Câteva exemple.

(1) N_{max}, Z_{max}, Q_{max}, R_{max}: multimile N, Z, Q respectiv R (de numere naturale, întregi, rationale, reale) completate cu elementul $-\infty$ cu rol de "zero" (ε), cu operatiile $\oplus = \max$, $\otimes = +$, cu 0 ca element "identitate"

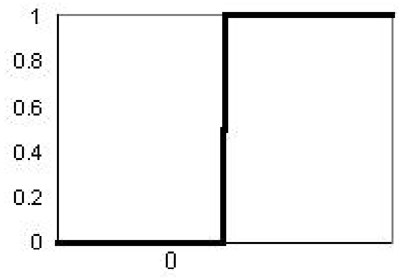
(2) Z_{min}: multimea Z completată cu $+\infty$ si cu operatiile $\oplus = \min$, $\otimes = +$. Elemnetul "zero" este $+\infty$, elementul "identitate" este 0.

În formulele prezentate, algebra în care se fac calculele este aceea a dioidului R_{max}.

METODE NECONVENTIONALE ÎN MODELAREA SI SIMULAREA SISTEMELOR DE PRODUCTIE

Retele neuronale

O descriere fie si sumară a unui neuron natural nu poate ocoli caracterul lui de dispozitiv de calcul cu un număr de intrări (*stimuli*) si o iesire unică pe un *axon*, iesire care uzual există sau nu există (este nulă) si care poate servi ca intrare pentru un alt neuron sau, în general, pentru alti neuroni în cadrul unor asa-zise *sinapse*.
Intrările unui neuron sunt combinate într-un gen de intrare unică prin însumare ponderată după anumite reguli a intrărilor propriu-zise. Există un prag de sensibiltate sub care iesirea neuronului este nulă. Depăsirea pragului produce o iesire, mereu aceeasi, asadar neuronul este principial un element cu iesire binară, pe numai două niveluri. În general, functia care leagă iesirea neuronului de intrarea lui sintetică se numeste functie de activare. Functia de activare tip prag asociată uzual cu neuronii naturali este ilustrată în figura alăturată. Se observă pragul de sensibilitate nenul si se sugerează o tranzitie în timp finit de la o stare, cea cu iesire nulă, la cealaltă stare cu iesire nenulă. Ca deobicei, nici în cazul neuronilor nu este posibilă variatia instantanee a unei mărimi, a iesirii lui.

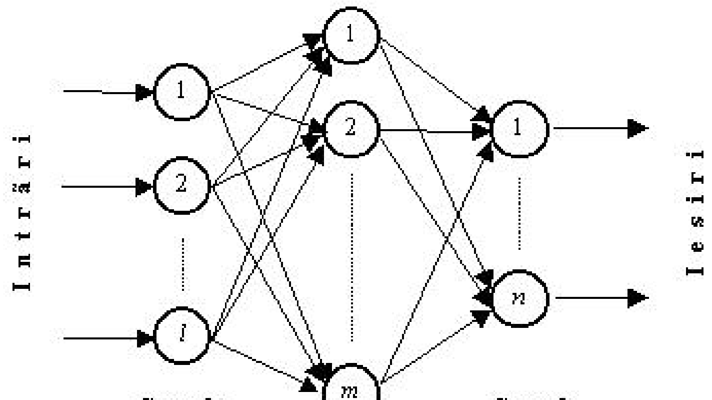


Posibilitatea de interconectare a neuronilor este foarte diversă si de aici structurile foarte variate si complexe ale sistemelor si subsistemelor nervoase precum si capacitatea lor de a executa calcule paralele de mare amploare. Sistemele neuronale au în plus capacitatea de a învăta. Toate aceste caracteristici au atras atentia de timpuriu tehnicienilor în încercarea lor de a

simula prin elemente de calcul procesele inteligente care au loc în sistemele nervoase, deseori în legătură directă cu sisteme tehnice foarte concrete.
În contiunare discutia se limitează la retelele neuronale organizate în straturi, adică fără cicluri în care un neuron ar putea furniza siesi, pe o cale mai mult sau mai putin ocolită, intrări.

Retele neuronale artificiale stratificate

Într-o retea neuronală artificială stratificată unitătile neuronale procesoare de informatie sunt dispuse într-o secventă de trei sau mai multe straturi de neuroni. Iesirile neuronilor dintr-un strat ponderate convenabil sunt intrări pentru neuronii care apartin exclusiv stratului următor sau sunt iesiri ale retelei dacă este vorba de stratul de iesire. Primul strat primeste intrările (stimulii) din ambiantă. Ultimul strat produce iesirile, în fond rezultatul unui calcul mai mult sau mai putin complex. Intrările neuronilor din straturile interioare, ascunse si ale ultimului strat, cel de iesire sunt combinatii liniare ale iesirilor produse de neuronii din stratul premergător. Coeficientii acelor combinatii liniare sunt numite *ponderi* si au un rol foarte importatnt în asa-zisa *instruire* a unei retele neuronale, într-un *proces de învătare* care face o structură cu neuroni stratificată să fie adaptată unui anumit scop tehnic sau tehnologic. Rolul oricărui strat neuronal interior, ascuns este acela de a reformula si de a reaplica iesirile stratului anterior pentru a obtine o reprezentare mai capabilă a separa, a clasifica datele de la intrarea retelei. Straturile interioare permit atasarea unei semantici particulare combinatiilor de intrări ale retelei.



Structura retelelor neuronale stratificate poate fi foarte diferită dacă se iau în considerare numărul de straturi si numărul de neuroni în fiecare strat. Figura de mai sus arată structura unei retele neuronale cu trei straturi, unul de intrare, unul ascuns si unul de iesire, cu *l*, *m*, respectiv *n* celule neuronale. Trebuie spus că stratul de intrare al unei retele neuronale artificiale are uzual numai rolul de a

pregăti intrările stratului următor. Neuronii din primul strat au câte o singură intrare pe care o aduc prin translatie si prin scalare la valori potrivite pentru a fi stimuli valabili pentru celulele neuronale din stratul următor.
Un răspuns la întrebările posibile si legitime referitoare la structurarea unei retele neuronale artificiale a fost dat cu multă vreme în urmă de matematicianul Kolmogorov în cadrul teoriei aproximării functiilor. Astfel, fiind dată o functie continuă $\phi : I^d \rightarrow R^c, \phi(x) = y$, unde $I = [0, 1]$ si, în consecintă, I^d este cubul unitate d-dimensional, functia ϕ poate fi implementată într-o retea neuronală cu exact trei straturi, cu d unităti în stratul de intrare, cu $(2d + 1)$ neuroni în unicul strat ascuns si cu c unităti în stratul de iesire.
Stratul ascuns, interior realizează aplicatia

$$z_k = \sum_{j=1}^{d} \lambda^k \psi(x_j + \varepsilon k) + k$$

în care x_j sunt intrările retelei, λ o constantă reală si ψ o functie, ambele independente de functia de reprezentat ϕ, iar ε este un număr rational pozitiv, mărginit. Functia ψ de activare a neuronilor din stratul ascuns trebuie să îndeplinească cunoscuta conditie a lui Lipschitz $|\psi(u) - \psi(v)| \le c|u - v|^\alpha$ pentru orice $\alpha \in (0,1]$ si pentru orice argumente $u, v \in I^d$.
Stratul de iesire face aplicatia

$$y_i = \sum_{k=1}^{2d+1} g_i(z_k)$$

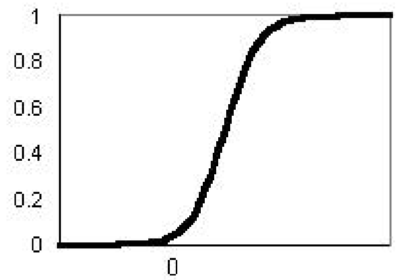
unde functiile g_i, $i = 1, 2, ..., c$ sunt reale si continue si depind de ϕ si ε.
Teorema dată de Kolmogorov este numai o teoremă de existentă. Construirea efectivă a functiilor ψ si g_i este deschisă. Posibilitătile de aproximare a functiei $\phi(x)$ cu functii de un gen sau altul rămâne de discutat în continuare.
În procesul de instruire/învătare pentru retelele neuronale artificiale stratificate se utilizează o multime de învătare H alcătuită din perechi (i^k, t^k), $k = 1, 2, ..., N$ de vectori de intrare (i de la *input* – intrare) si de vectori de răspuns asociati (t de la *target* – tintă) cu valori observate experimental. Operatiile de bază cunoscute si sub numele cuprinzător de *regula delta generalizată* sunt:

- Se aplică retelei neuronale intrări (stimuli) i^k din multimea de învătare;
- Se calculează pas cu pas iesirile tuturor unitătilor retelei neuronale, având în vedere functiile de activare specifice, în cele din urmă iesirile o^k (o de la *output* – iesire);
- Se compară pe baza unui criteriu prestabilit vectorul o^k (iesire a stratului ultim al retelei, stratul de iesire) cu vectorul de iesire t^k pereche în multimea H cu intrarea aplicată retelei;
- Se calculează eroarea si se propagă măsura ei în sens invers, de la iesire către intrare;
- Se încearcă minimizarea erorii la fiecare etapă prin modificarea ponderilor retelei.

Pentru minimizarea erorii E de predictie a iesirilor t^k prin iesirile calculate o^k se poate utiliza orice metodă de determinare a extremelor unei functii, în cazul în discutie functia care măsoară eroarea de predictie. Metodele de gradient sunt desigur utilizabile dacă functia care măsoară diferentele (în sens larg) între t^k si o^k este derivabilă. Vectorul derivatelor partiale $\partial E / \partial w_{ji}$ în raport cu ponderile w_{ij} atasate intrărilor pentru celula j din stratul i dă directia de modificare a ponderilor, care trebuie să fie în sensul invers al vectorului gradient. Asadar, modificările Δw_{ij} trebuie să fie proportionale cu componentele vectorului gradient cu semn schimbat. Calculul acestor derivate contine o procedură de derivare a unor functii care la rândul lor au ca argumente alte functii. Intervin aici inevitabil functiile de activare ale celulelor neuronale. Dacă acestea sunt de tipul prag teoretic (salt instantaneu), derivata lor este pretutindeni nulă, iar în punctul corespunzător pragului derivata nu există. Pentru a evita acest inconvenient, pragul teoretic – salt pentru argument egal cu pragul de sensibilitate al neuronului – este înlocuit în aplicatii de functia sigmoidală care are expresia

$$\sigma(x) = \frac{1}{1 + e^{-\alpha x}}$$

si înfătisarea din figura alăturată.



Din coeficientul pozitiv α se poate aranja ca panta de trecere de la nivelul minim la cel maxim (si invers) să fie oricât de abruptă: cu cât mai mare α cu atât mai mare panta si, la limită, când α este foarte mare, sigmoida devine pragul ideal. Avantajul functiei sigmoidale este acela că ea este derivabilă pretutindeni, asadar metodele de minimizare a distantei dintre iesirile prezise si cele observate, bazate pe gradient sunt deplin abordabile.

Retelele neuronale artificiale sunt deja larg utilizate pentru a rezolva probleme de învătare în diverse domenii. Prin utilizarea unor date experimentale existente, retelele neuronale *învată* în fond relatiile între intrări si iesiri.

Relatiile neliniare sunt cu totul empirice si nu sunt bazate pe vreo teorie din fundamentele fizicii etc. Sub acest unghi, retelele neuronale sunt pur si simplu modele regresionale complexe a căror structură este determinată empiric. Desi retelele neuronale artificiale sunt inspirate de retelele de celule nervoase ale organismelor vii, dezvoltările aplicative ulterioare, până la cele mai evoluate ale acestor retele numite si modele conexioniste sunt puternic influentate de dezvoltările recente înregistrate de analiza functională.

În domeniul ingineriei sistemelor, inclusiv al celor economice, se observă cu certitudine o explozie a interesului academic dar si industrial-comercial fată de retelele neuronale artificiale cu aplicatii în proiectarea de procese si de produse, în operarea si reglarea automată a proceselor, multe din ele de remarcabilă complexitate. Câteva exemple:

- Generarea de modele neliniare destinate proiectării sistemelor de reglare predictivă, fixe sau adaptive
- Diagnoza functionării defectuoase a sistemelor si identificarea cauzelor
- Monitorizarea si interpretarea tendintelor proceselor continue si/sau discontinue, cu evaluarea performantelor tehnologice si a calitătii produselor
- Modelarea comportării haotice a sistemelor dinamice deterministe.

Varietatea de reprezentări pe care retelele neuronale le pot cuprinde (booleene, calitative, semicantitative si/sau analitice/cantitative), gradul mare de paralelism al calculelor pe care retelele îl pemit si simplitatea structurii lor le-au transformat în instrumente de mare popularitate printre ingineri, cu utilizări pentru rezolvarea unei varietăti largi de probleme.

O retea neuronală tipică (din cele stratificate, deocamdată cele mai utilizate) este constituită din mai multe straturi de noduri interconectate, fiecare cu o functie de activare si ponderi pe fiecare arc care conectează nodurile retelei între ele. Iesirea fiecărui nod este o functie neliniară de toate intrările sale. Astfel, reteaua este o dezvoltare a relatiei neliniare necunoscute între intrările x si iesirile F într-un spatiu generat de asa-numitele functii de activare ale nodurilor retelei. În particular, învătarea prin propagare directă în retele stratificate poate fi privită ca sintetizarea unei aproximări a unei functii multidimensionale în spatiul generat de functiile de activare $\phi_i(x)$, $i = 1, 2, ..., m$, adică

$$F(x) = \sum_{i=1}^{m} c_i \phi_i(x)$$

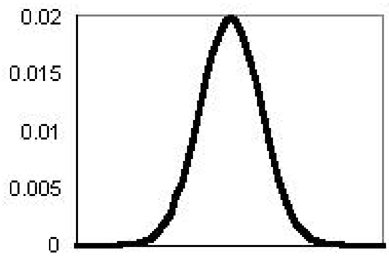
Cu date empirice la dispozitie, cu functiile de activare date si cu o topologie a retelei cunoscută, parametrii c_i, $i = 1, 2, ..., m$ sunt ajustati astfel încât eroarea aproximării să fie oricât de mică.

S-au prezentat mai devreme două functii de activare, functia prag ideal si functia sigmoidală. Ambele au răspunsuri pentru orice intrare, nu importă cât de mare sau cât de mică este acea intrare. De aceea ele sunt calificate drept functii de activare globale si nu sunt singurele în genul lor. Ele sunt doar cele mai cunoscute, prima utilizată pentru celulele neuronale din retelele numite si perceptroni si cealaltă utilizată pe larg în retelele stratificate cu învătare prin propagare secventială inversă (BPN – *BackPropagation Network*). Asadar, în general, neuronii cu functii de activare globale sunt activi pe un domeniu larg de valori ale intrărilor si asigură o aproximare globală a datelor empirice.

Cu functii de activare sigmoidale, cu retele neuronale de tipul stratificat, secvential cu un singur strat ascuns compus din m noduri, se pot aproxima functii foarte diverse prin functii din multimea

$$S_m \equiv \left\{ f(x) / f(x) = \sum_{i=1}^{n} c_i \sigma(x w_i + \theta_i), w_i \in R^d, c_i, \theta_i \in R \right\}$$

unde w_i, c_i, θ_i sunt parametri ajustabili. Se poate arăta că dacă m este suficient de mare atunci orice functie continuă poate fi aproximată oricât de exact conform cu formula de mai sus.
O alternativă la functiile de activare globale o constituie functiile de activare locale. Acestea produc iesiri ale neuronului nenule cu precădere într-o vecinătate restrânsă a unor valori de intrare. Iesirea lor se estompează pentru valori situate departe de centrul de răspuns maxim al functiei de activare si, implicit, de centrul de maximă receptivitate a celulei neuronale căreia functia îi este atasată.
Functiile de tipul radial de pildă sunt în esentă locale si sunt utilizate în retelele cu baze de functii radiale (RBFN - *Radial Basis Function Network*). Figura care urmează reprezintă o asemenea functie, functia gaussiană.



În general, o functie radială este o functie de o normă a diferentei dintre intrarea efectivă x a celulei si intrarea x_i care maximizează iesirea acelei celule

$$\phi_i(x) = h(\|x - x_i\|)$$

si este asociată unui nod sau centru de coordonate x_i.
Functia gaussiană în varianta ei multidimensională

$$\phi_i(x) = \frac{\sqrt{\det W}}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}(x - x_i)^T W (x - x_i) \right), \; x \in R^n$$

cu W o matrice pozitiv definită (o matrice de ponderi în directii diverse din R^n) este de tipul radial.
În cazul unidimensional ilustrat putin mai devreme, aceeasi functie se scrie sub forma

$$\phi_i(x) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{(x - x_i)^2}{2\sigma_i^2} \right), \; x \in R$$

Retelele de tipul RBFN pot si ele să aproximeze functiile continue cu o eroare oricât de mică. Retelele de acest tip necesită o prealabilă sortare a intrărilor, o operatie de aglomerare (*clustering*) în clase de intrări similare.

Proceduri de instruire pentru retelele neuronale cu baze de functii radiale

Proiectarea unei retele neuronale cu bază de functii radiale implică determinarea parametrilor c_k , t_k si σ_k pentru fiecare celulă neuronală în parte, care fac cât de mică posibil eroarea globală de aproximare. Parametrii t_k si σ_k exprimă coordonatele punctelor de maximă receptivitate a celulei neuronale ascunse k, respectiv aria de sensibilitate acoperită de acea celulă a cărei functie de activare este de tip radial. Parametrul c_k este coeficientul din formula de interpolare implementată prin structura de neuroni proiectată.

Problema stabilirii celor trei parametri pentru fiecare din celulele retelei neuronale poate fi rezolvată ca o singură problemă de optimizare globală prin instruire supravegheată, cu alte cuvinte pe baza unei multimi de învătare. Se poate însă proceda si la o determinarea etapizată. Într-o primă atapă se stabilesc centrele t_k si deviatiile standard σ_k în mod nesupervizat, pe măsură ce se acumulează date experimentale. În a doua etapă se stabilesc coeficientii c_k printr-o procedură de optimizare prin instruire supravegheată. Această procedură în două faze este, se pare, mai eficientă. Iat-o descrisă sumar mai departe.

Faza I. Instruire pentru autoorganizare. În această fază se calculează centrele t_k ale celor K functii de bază radiale si extinderea lor dată de σ_k. Pentru a găsi cele K centre de maximă receptivitate din setul de intrări al exemplelor de instruire se foloseste un algoritm standard de aglomerare cu k medii (*k-means clustering algorithm*). Fiecare grupare, aglomerare (*cluster*) se leagă de un nod ascuns al retelei. Centrul grupării determină valoarea t_k a functiei radiale din bază. Pasul curent alocă noduri numai pentru regiunile unde există date. Lărgimea (sau dispersia) fiecărui câmp este apoi stabilită printr-o euristică a contiguitătii. Multe euristici de tipul vecinului celui mai p-apropiat (*p-nearest neighbor*) pot fi utilizate. De exemplu, lărgimea poate fi dată de media geometrică $\sigma = \sqrt{d_1 d_2}$ unde d_1 si d_2 sunt distantele euclidiene de la centrul k la două centre cele mai apropiate. Această euristică asigură o oarecare suprapunere pentru fiecare unitate cu unităti vecine ceea ce conduce la o interpolare netedă pe spatiul intrărilor. Instruirea autoorganizantă din faza de fată reduce efortul de instruire supervizată din faza a doua în care trebuie evaluate numai ponderile.

Faza II. Minimizarea erorii medii pătratice. Ponderile c_k ale functiilor de bază radiale sunt găsite prin minimizarea erorii medii pătratice

$$E = \sum_k [y_k - F(x_k)]^2$$

Determinarea ponderilor este o problemă liniară a cărei convergentă este aproape sigură. S-au sugerat multe îmbunătătiri si alternative pentru instruirea retelelor cu baze de functii radiale. Unii autori si cercetători au utilizat coeficienti care sunt functii liniare de intrări, care permit functii gaussiene asimetrice si o functie pătratică de cost de forma

$$E = \left[\frac{1}{2}\sum_i y_i^2\right] - \sum_k c_k\left[\sum_i y_i \Phi_k(x_i)\right] + \frac{1}{2}\sum_k\sum_l c_k c_l\left[\sum_i \Phi_k(x_i)\Phi_l(x_i)\right]$$

cu (x_i , y_i) perechi intrare-iesire si cu $\Phi_k(x_i)$, $\Phi_l(x_i)$ functii din bază. Metoda aceasta se arată a fi superioară precedentei la predictia seriilor de timp haotice. Faza primă, euristică, de instruire a retelelor cu baze de functii radiale necesită reluări cu diferite numere de unităti ascunse pentru a obtine un optim structural al retelei.

În cuprinsul acestei sectiuni s-a făcut referire la stabilirea extremelor unor functii care exprimă distanta între observatii experimentale si iesirile calculate ale unui model cum este de fapt o retea neuronală. Metodele de stabilire a acestor extreme, parte a procesului de instruire au utilizări mai largi, în împrejurări variate. De aceea, sectiunea următoare este consacrată descrierii (uneori sumare a) acestor metode.

Metode de stabilire a unor solutii optime

Frecvent, în rezolvarea unor probleme de naturi foarte diferite inginerii au de ales între mai multe solutii posibile si fezabile. Alegerea nu se face niciodată la întâmplare ci pe baza unor criterii care presupun o optimizare.

Câteva exemple de optimizări au fost deja parcurse în alte capitole ale acestei lucrări ca si în cadrul lucrărilor aplicative de la această disciplină, *Modelarea si simularea sistemelor de productie*: optimizări în cadru liniar, rezolvate prin metodele programării liniare, optimizări pe grafice-retea etc. S-a retinut încă de atunci, dacă faptul nu era cunoscut deja, că există o functie obiectiv care trebuie maximizată/minimizată si un număr de variabile de decizie prin modificarea cărora se obtine extremul urmărit dacă acesta există. Este de rememorat totodată faptul că variabilele de decizie trebuiau să satisfacă un număr de restrictii. Fie si numai din acele exemple relativ simple parcurse în capitolele respective sau la lucrări se poate extrage forma generală a unei probleme de optimizare, componentele unei astfel de probleme:

- Functie obiectiv
- Variabile de decizie
- Restrictii
- Algoritm de stabilire a extremelor

Varietatea mare de probleme de optimizare provine din:

- Particularitătile functiei obiectiv: liniaritate (neliniaritate), multime de valori compactă sau discretă, continuitate, derivabilitate, uni- sau multimodalitate.
- Numărul variabilelor de decizie si tipul lor

- Număr de restrictii
- Caracterul determinist sau aleator al problemei

La capitolul algoritmi de optimizare varietatea este la fel de mare. Fără pretentie de exhaustivitate se pot enumera:

- Algoritmi cu evaluare directă
- Algoritmi bazati pe gradient
- Algoritmi de căutare aleatoare
- Algoritmi genetici

Desigur, există algoritmi hibrizi, adică algoritmi de un gen din cele mentionate "contaminati" cu elemente specifice algoritmilor de alte genuri.

În continuare se consideră functii obiectiv de forma generală $f(x_1, x_2, \ldots, x_n)$ care pot include si anumite "penalităti" la apropierea de vreuna dintre restrictii.

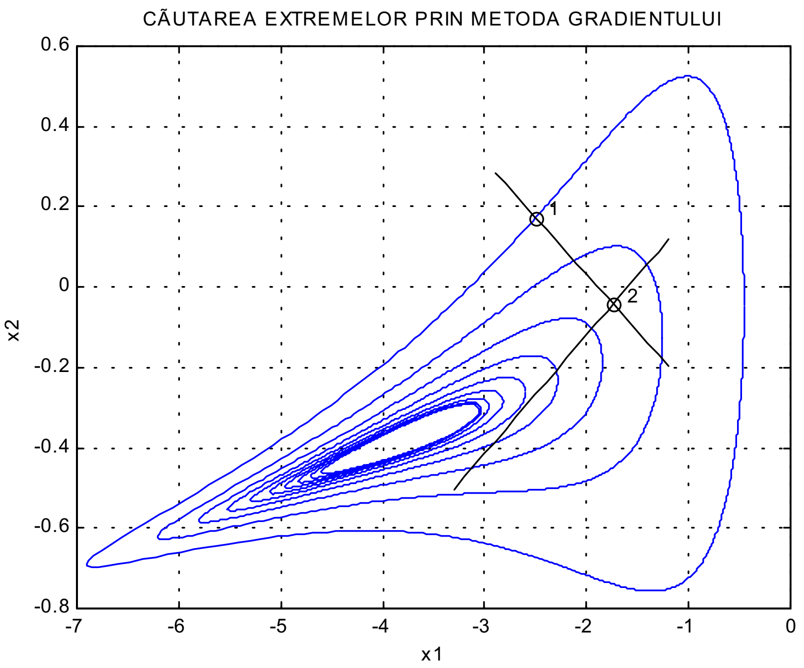
Algoritmii cu evaluare directă constau în evaluarea functiei obiectiv într-un număr de puncte din spatiul variabilelor de decizie, denumit si spatiu de căutare (a optimului). Este o metodă care se aplică la probleme cu dimensionalitate redusă: spatiul de căutare cu maximum 2-3 dimensiuni. Din "ploaia" de evaluări, de regulă sistematică, se retine solutia cea mai favorabilă. Eficienta metodei este discutabilă chiar la dimensiunile mentionate: consum de timp de calcul uneori mare, stabilirea optimului cu o precizie de cele mai multe ori îndoielnică. Are avantajul că nu cere calităti speciale ale functiei obiectiv (continuitate, derivabiliate etc.)

Algoritmii bazati pe gradient se aplică exclusiv în cazul functiilor obiectiv derivabile în raport cu fiecare dintre variabilele de decizie. Gradientul într-un punct din spatiul de căutare este vectorul de valori ale derivatelor partiale ale functiei obiectiv în acel punct (indicele superior T pentru operatia de transpunere).

$$grad\, f = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}^T$$

Directia lui indică directia în care functia obiectiv are cea mai rapidă variatie. O deplasare în sensul vectorului gradient (deplasare care se obtine prin modificări ale variabilelor de decizie proportionale cu valorile derivatelor componente ale gradientului) produce o crestere a functiei obiectiv. O deplasare în sens invers produce o scădere a functiei obiectiv. Desigur, evaluarea derivatelor partiale consumă timp dar pasii pe directia gradientului duc de cele mai multe ori la îmbunătătiri rapide ale functiei obiectiv. Se practică adesea proceduri de accelerare a deplasării pe directia respectivă, dacă îmbunătătirile sunt promitătoare, sau de decelerare, dacă îmbunătătirile s-au plafonat sau au devenit înrăutătiri. Asadar, gradientul nu se redefineste prin calcul după fiecare evaluare a functiei obiectiv ci numai după ce căutarea pe directia gradientului încetează a mai fi productivă, aducătoare de valori mai bune pentru functia al cărui extrem se caută, maxim sau minim, de la caz la caz.

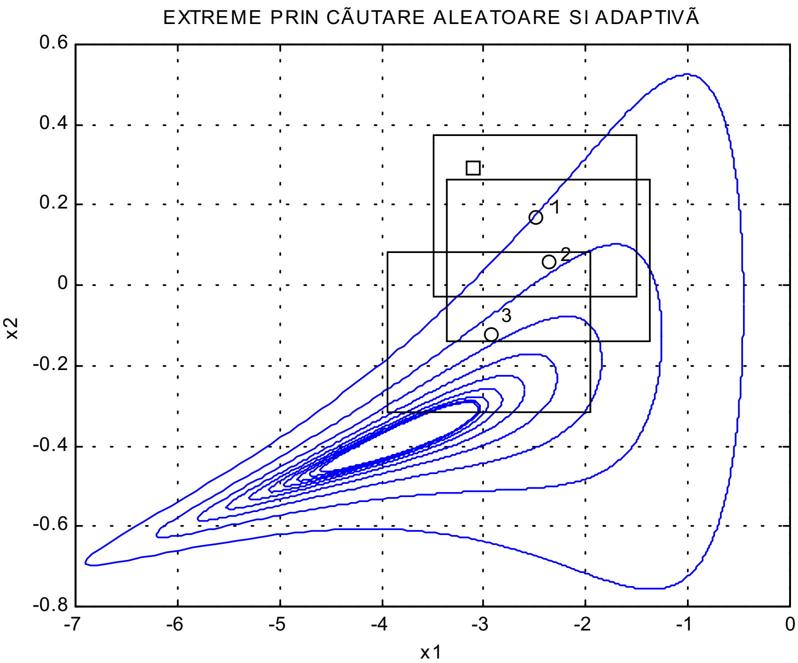
Printr-o similitudine cartografică s-a reprezentat alăturat o functie obiectiv de două variabile, x_1 si x_2, prin curbe de nivel, locuri geometrice alcătuite din puncte în care functia obiectiv ia aceleasi valori pentru multiple perechi de valori (x_1, x_2). Curbele centrale sunt din ce în ce mai apropiate de extrem, curba periferică este cea mai slabă prin prisma valorilor functiei. Sunt reprezentate două directii ale gradientului. Una este pentru gradientul evaluat în punctul 1, punct de initiere a căutării. Se observă că directia de cea mai rapidă variatie a functiei nu poate fi decât transversală fată de curba de nivel care trece prin punctul respectiv. Ea este chiar perpendiculară pe tangenta la curbă: tangenta la curba de nivel este o directie în care functia are variatie nulă (functia este constantă, cel putin local). Un număr de evaluări ale functiei în puncte situate pe directia gradientului aduce mai întâi o îmbunătătire a valorilor ei, apoi o înrăutătire. Punctul 2 din figură este ultimul punct bun de pe directia gradientului evaluat în punctul 1. Aici se reevaluează gradientul si se stabileste o nouă directie de căutare, reprezentată si ea în figură. Procedura se repetă până când se atinge extremul căutat. Desigur, o reprezentare similară pentru functii de mai multe variabile nu este posibilă dar principiile căutării si algoritmul rămân.



Metoda gradientului si numeroasele ei variante nu sunt totdeauna conducătoare către optimul functiei obiectiv. Dacă functia este multimodală, adică are mai multe extreme, căutarea se poate opri într-un extrem local, îndepărtat de optim. Tot printr-o similitudine topografică/cartografică, un relief ondulat poate cuprinde mai multe înăltimi si mai multe văi închise (căldări) sau deschise. Căutarea unui maxim de altitudine poate esua într-un vârf care nu este cel mai înalt în peisaj. Căutarea unei cote minime se poate încheia într-o căldare care nu-i cea mai adâncă în regiunea explorată. Metodele de gradient au de asemenea dificultăti în căutarea eficientă când functia obiectiv are variatii rapide similare unor văi adâci si abrupte, ca într-un relief cu ravene. Dimensionalitatea mare a spatiului de căutare reduce aficienta căutării extremelor prin metodele care se bazează pe evaluarea gradientului.

Metodele aleatoare de căutare prezintă o protectie mai bună la extremele multiple, la multimodalitatea functiilor obiectiv al optimizării. Mai au avantajul că merg si în cazurile în care functia obiectiv nu este derivabilă. Este în fond o metodă de evaluare directă numai că amplasarea punctelor în care functia obiectiv este calculată din nou si din nou este aleatoare.

Figura alăturată ilustrează o căutare aleatoare a extremului aceleiasi functii despre care s-a discutat si la metoda gradientului. Metoda este aleatoare dar este si adaptivă. Ce înseamnă adaptivitatea se întelege mai bine dacă se retine faptul că orice căutare de extrem nu se întinde niciodată pe spatii nelimitate: spatiul de căutare este fatalmente finit. Finitudinea lui poate exclude însă zona unde se află alte extreme, poate chiar extremul-optimul căutat.



În exemplul din figură căutarea începe prin evaluări ale functiei în puncte “semănate” aleator în interiorul dreptunghiului cu centrul de simetrie în punctul 1. Evaluările pot duce la valori ale functiei mai slabe decât cea din centrul dreptunghiului (un exemplu este punctul marcat cu un pătrat). Primul puct mai bun decât punctul 1 (marcat aici cu 2) este retinut si dreptunghiul se deplasează prin translatie astfel încât acest punct să devină centrul lui de simetrie. Explorarea aleatoare continuă în domeniul delimitat de acest dreptunghi în noua lui pozitie. Vor fi aproape sigur câteva valori care nu corespund (punctele respective nu sunt reprezentate), dar va apărea si în acest caz un punct mai bun:

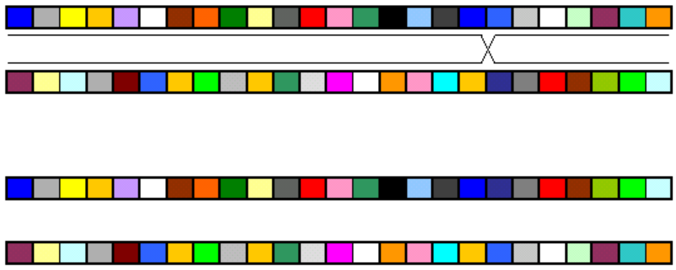
punctul 3. O nouă deplasare a dreptunghiului, centrată de data aceasta pe cel mai recent punct bun aduce – se observă (ceea ce la o functie de mai mult de două variabile din păcate este imposibil) – o zonă bogată în puncte mai bune decât tot ce s-a obtinut până acum. Perspectiva îmbunătătirii functiei obiectiv creste evident. Adaptare căutării se face prin această deplasare. Acum, dacă presupunem că dreptunghiul migrator atinge fie si numai cu un vârf vecinătatea unui alt extrem (local) există sansa ca operatia de căutare să fie orientată către acel extrem care poate fi mai bun decât alte extreme. Asta este o protectie la ignorarea unor extreme multiple pe care metodele de gradient nu o au.
Sunt împrejurări în care o combinare a metodelor de gradient cu cele de căutare aleatoare aduce o însumare a calitătilor celor două metode. Desigur, functiile ale căror extreme se caută trebuie să fie derivabile. În această situatie o căutare bazată pe gradient, când dă semne de stationaritate este oprită si o căutare aleatoare, uneori grosieră oferă sansa unei iesiri dintr-un extrem local prin "nimerirea" vecinătătii unui alt extrem mai bun decât cel localizat prin utilizarea gradientului.
Pentru problemele cu dimensionalitate foarte extinsă, acesta este cazul instruirii unei rete neuronale unde varabilele de decizie sunt ponderile, se recurge la metode împrumutate de la regnul viu. Sectiunea imediat următoare contine un asemenea recurs.

Algoritmi genetici

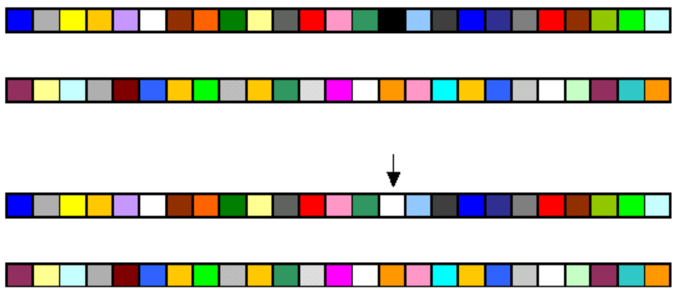
Problemele ingineresti cu dimensionalitate mare sau foarte mare se pot trata prin metode bazate pe algoritmii genetici. Stabilirea extremelor unor functii multimodale, structurarea optimă si instruirea retelelor neuronale sunt exemple de asemenea probleme. Algoritmii genetici sunt un împrumut din biologie si se bazează pe evolutionismul darwinian.
Se consideră o populatie alcătuită din indivizi descrisi de structuri numite cromozomi. Cromozomii sunt uzual structuri liniare, ansambluri de gene. Figura alăturată ilustrează doi indivizi prin cromozomii lor, genele fiind reprezentate prin culori.

Orice populatie este în evolutie. Indivizii care o alcătuiesc se combină în perechi pentru a genera urmasi. Procedeul curent este cel al combinării-încrucisării. Prin combinare rezultă descendenti care sunt la rândul lor caracterizati de cromozomi. Cromozomii lor rezultă printr-o lectură încrucisată a cromozomilor parentali, în linii mari conform schemei din figura care urmează. În partea de jos sunt reprezentati prin cromozomii specifici descendentii rezultati.

Nu este obligatoriu ca din combinare să rezulte doi descendenti dar în multe aplicatii tehnice aplicarea *operatorului de combinare* produce doi descendenti. Desigur, punctul de comutare a lecturii de la un cromozom la celălalt poate fi pozitionat si altundeva. De asemenea, pot exista si mai multe puncte de traversare.



Lectura cromozomilor parentali se poate face corect dar se poate face si cu eroare. Dacă s-a produs o eroare, se spune că a avut loc o mutatie. Asadar, există un al doilea operator genetic, *operatorul de mutatie*. Figura următoare ilustrează efectul unei mutatii. Sunt prezentati din nou descendentii rezultati prin lectura corectă a genelor, apoi descendentii dintre care unul este afectat de o mutatie la gena marcată cu săgeată.



În aplicatiile ingineresti se vorbeste de populatii de solutii ale unei probleme si de determinarea evolutivă a solutiei acelei probleme. Este vorba mai ales de probleme complexe, de dimensionalitate excesivă pentru care nu sunt căi analitice de solutionare, iar enumerarea tuturor solutiilor acceptabile este o iluzie. Si aici, ca si în cazul populatiilor biologice se vorbeste de adecvarea mai bună sau mai slabă a solutiilor la problema tratată, întocmai cum indivizii unei specii sunt adecvati mai mult sau mai putin la problema supravietuirii într-un mediu generator de variate provocări. Si într-un caz si în altul principiul darwinian al selectiei naturale "*supravietuiesc cei mai adecvati*" lucrează sistematic pentru adaptarea solutiilor la problema fomulată, respectiv a indivizilor la problema supravietuirii si implicit a perpetuării.
Din expunerea generală de mai sus rezultă că problemele tehnice si economice se pot rezolva evolutiv dacă există o codare prin cromozomi adecvati a solutiilor admisibile si dacă se defineste corespunzător o functie de adecvare.

Cromozomii din aplicatiile ingineresti pot avea forme diverse. La fel functiile de adecvare. Cea mai frecventă codare este cea binară: cromozomii sunt siruri de biti, genele sunt bitii însisi.
Orice formă ar avea cromozomii, solutionarea unei probleme prin utilizarea algoritmilor genetici parcurge o cale evolutivă, solutia se obtine prin *evolutie*. Algoritmul porneste de la o *populatie* de solutii reprezentate prin *cromozomi*. Solutiile dintr-o populatie sunt utilizate pentru a forma o nouă populatie de solutii. Motivatia este cât se poate de naturală: speranta că noua populatie va fi mai bună decât populatia veche. Solutiile alese pentru a produce solutii noi, pentru a produce *descendenti*, sunt alese pe baza potrivirii lor cu mediul problemei de solutionat: cu cât sunt mai adecvate, cu atât ele au mai mari sanse de a se *reproduce*.
Procedura este repetată până când s-a generat un număr dat de populatii succesive sau o anumită conditie de adecvare a fost atinsă.
Algoritmii genetici (AG) cuprind în general pasii următori:

1. Generarea aleatoare a unei populatii initiale de *n* solutii acceptabile ale problemei, reprezentate de *n* cromozomi
2. Evaluarea unei functii de adecvare *f*(*x*) pentru fiecare cromozom *x* din populatie
3. Crearea unei populatii noi prin repetarea pasilor următori până ce populatia nouă este completă
    a. Selectia: se selectează o pereche de cromozomi părinti în acord cu adecvarea lor (cu cât sunt mai adecvati cu atât au sanse mai mari de a fi alesi pentru reproducere)
    b. Încrucisarea: cu o probabilitate de încrucisare dată se încrucisează părintii pentru a genera o pereche de descendenti (dacă nu are loc o încrucisare descendentii vor fi cópii identice ale părintilor)
    c. Mutatia: cu o probabilitate precizată se modifică unele pozitii, unele gene din cromozomii descendentilor
4. Populatia generată înlocuieste populatia veche si este folosită pentru o nouă parcugere etapă cu etapă a algortimului
5. Dacă conditia de oprire este atinsă, algoritmul se încheie si se retine solutia cea mai bună din populatia curentă, care este si ultima
6. Dacă conditia de oprire nu este atinsă se reiau evaluările de la pasul **2.**

Liniile generale ale algoritmilor genetici date mai sus au implementări variate.
Una din probleme este, asa cum s-a spus, cum să se creeze cromozomii, cum să se realizeze această codare a indivizilor dintr-o populatie. În functie de forma cromozomilor se definesc cei doi operatori de bază ai algoritmilor genetici, combinarea-încrucisarea si mutatia.
O altă problemă este selectarea judicioasă a părintilor pentru încrucisare. Selectarea se poate face în moduri diferite dar ideea generală este a retine părintii dintre cei mai buni, în speranta că descendentii lor vor fi si mai buni. Poate interveni un dubiu si anume că alcătuirea populatiei noi numai din descendenti ar putea conduce la pierderea cromozomilor cei mai buni din generatia precedentă. Asta se poate întâmpla si, de aceea, se foloseste uneori

asa-zisul *elitism*. Asta înseamnă că cel putin una din cele mai bune solutii din generatia curentă este retinută prin copiere în generatia următoare ceea ce o face viabilă poate până în faza finală a evaluărilor.
Modul cel mai obisnuit de codare cromozomică constă în constituirea unei secvente de valori binare.
Cromozomii arată în acest caz astfel:

Cromozomul *k*	1101100100110110
Cromozomul *l*	1101111000011110

Fiecare bit din secventă reprezintă o anumită caracteristică a solutiei. Uneori secventa poate reprezenta unul sau mai multe numere. Desigur, sunt si alte modalităti de codare. Codurile adoptate depind si de tipul problemei de rezolvat. Se pot coda, de pildă, direct numere întregi sau reale, uneori anumite permutări, structuri grafice etc.
Parametri pentru AG. Probabilitătile asociate încrucisării si mutatiei sunt parametri de bază ai algoritmilor genetici. Probabilitătile referitoare la încrucisări se asociază cu frecventa cu care un individ sau altul este selectat în vederea încrucisării: indivizii sau solutiile mai adecvate au probabilităti mai mari de a fi selectati pentru combinare, pentru aplicarea operatorului de încrucisare. Când punctul, altfel aleator, de comutare a lecturii de pe un cromozom pe celălalt este situat chiar pe prima sau pe ultima genă din secventa cromozomială descendentii sunt cópii identice ale părintilor. Încrucisarea este făcută în speranta cât se poate de naturală conform căreia cromozomii noi vor contine genele asociate părtilor bune din cromozomii parentali si acesti noi cromozomi vor reprezenta solutii mai bune ale problemei. Uneori se renuntă total la o generatie de solutii de îndată ce o nouă generatie este completă. Alteori este îngăduit ca o parte a populatiei să supravietuiască si în generatia următoare pentru a păstra solutiile cele mai perfectionate ca material genetic valoros pentru încrucisările efectuate în etapa/etapele viitoare.
La mecanismul încrucisărilor se recurge aproape în orice algoritm genetic cu o frecventă mare. Mutatia este folosită mai rar, mai curând ca accident. De aceea probabilitatea de aparitie a unei mutatii este fixată la valori mici, sub 0,1. Mutatia este folosită pentru a preveni stagnarea căutării într-o zonă de adecvare bună numai relativ la o vecinătate restrânsă, ceva analog unui extrem local în optimizare.
Un alt parametru important este dimensiunea populatiei mentinută de regulă constantă de la o generatie la următoarea. Dacă populatia este redusă, diversitatea cromozomială este modestă si algoritmul genetic are posibilităti slabe de încrucisare ceea ce se traduce în conducerea explorării pe un spatiu restrâns. Pe de altă parte populatiile prea numeroase fac ca algoritmii genetici să lucreze lent. O recomandare de luat în considerare are în vedere populatii de zeci de indivizi-solutii.

În una din lucrările aplicative prevăzute la disciplina *Modelarea si simularea sistemelor de productie* se propune spre studiu si observare actiunea de căutare a extremului unei functii de o variabilă cu foarte multe extreme, o functie multimodală a cărei expresie este

$$f(x) = -\frac{480}{4.2}\sin\frac{\pi x}{640}\left(1-\sin\frac{30\pi x}{640}\right)\left(1-\sin\frac{5.3\pi x}{640}\right)$$

Populatia initială este de 20 de solutii. Dimensiunea populatiilor următoare este aceeasi. Se practică elitismul total, adică la fiecare nouă generatie clasamentul adecvării solutiilor se întocmeste pe 40 de solutii vechi si noi. Sunt eliminate 20 de solutii din josul clasamentului indiferent dacă sunt printre ele solutii abia generate. Algoritmul genetic foloseste parametrii pe care observatorul îi poate stabili el însusi. Acestia sunt numărul de generatii propus pentru stoparea automată a algoritmului, apoi raportul, supraunitar desigur, între probabilitatea de selectare în vederea încrucisării a celei mai perfectionate solutii si a celei mai putin adecvate si în sfârsit, probabilitatea aparitiei unei mutatii.

B I B L I O G R A F I E

1. J.E.Beasley “OR-Notes”, http://mscmga.ms.ic.ac.uk/jeb/or/ , Imperial College, Londra, 2002
2. R.E.Bellman si S.E.Dreyfus “Programarea dinamicã aplicatã” Editura Tehnicã, Bucuresti, 1967
3. Gh.Boldur-Lãtescu, Gh.Ciobanu si I.Bãncilã “Analiza sistemelor complexe” Editura Stiintificã si Enciclopedicã, Bucuresti, 1982
4. S.Cãlin, Th.Popescu, B.Jora si V.Sima “Conducerea adaptivã si flexibilã a proceselor industriale” Bucuresti, Ed.Tehnicã 1988
5. G.Cohen “Théorie algébrique des systèmes à événiments discrets” Centre Automatique et Système, École des Mines de Paris, Fontainbleau & INRIA Rocquencourt, 1995
6. G.Cohen “Analysis y control de sistemas de eventos discretos: de redes Petri temporizadas al algebra” Universidad de Rosario, Argentina, 2001
7. S.E.Elmaghraby “Proiectarea sistemelor de productie” Editura Tehnicã, Bucuresti, 1968
8. A.Kauffmann “Metode si modele ale cercetãrii operationale” Editura Stiintificã, Bucuresti, 1967
9. L.Lasdon “Teoria optimizãrii sistemelor mari” Editura Tehnicã, Bucuresti, 1975
10. S.Lazãr “Analiza drumului critic” Editura Stiintificã, Bucuresti, 1968
11. O.Pãstrãvanu “Sisteme cu evenimente discrete. Tehnici calitative bazate pe formalismul retelelor Petri” Editura MATRIX-ROM, Bucuresti, 1997