Bogusław Guzik

PROGNOZOWANIE FUNKCJI

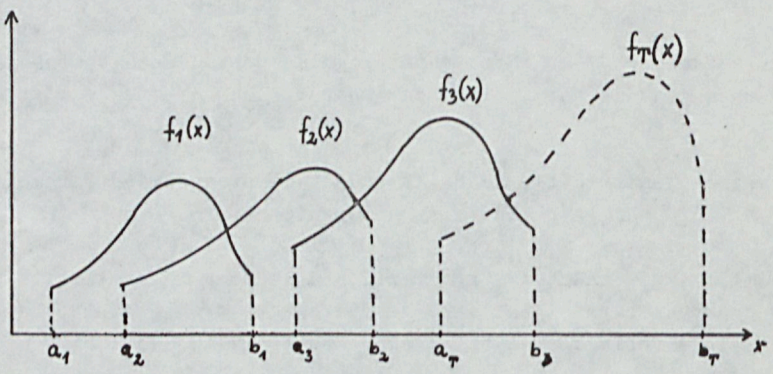
1. Wstęp

Niech dany będzie zbiór $y_1, y_2, \ldots, y_n, y_T$ wartości zmiennej y. Zakładamy, że z każdą wartością y_t $(t=1,2,\ldots,n,T)$ związano pewną funkcję $f_t(x)$ zmiennej rzeczywistej x. Dziedzinę funkcji $f_t(x)$ oznaczymy przez P_t; jest to przedział, gdy x jest zmienną ciągłą, np. przedział skończony $P_t = [a_t, b_t]$ lub zbiór, gdy x jest zmienną dyskretną. Funkcję i jej dziedzinę będziemy niekiedy nazywali kształtem i oznaczali symbolem $\{f_t(x), x \in P_t\}$.

Zadanie, jakie sobie stawiamy, to odgadnięcie kształtu $\{f_T(x), x \in P_T\}$ tzn. odgadnięcie funkcji $f_T(x)$ i jej dziedziny P_T na podstawie informacji o kształtach

$$\{f_t(x),\ x \in P_t\}\ ,\quad t=1,\ldots,n\ .$$

Ilustruje to rys. 1, na którym $f_1(x)$, $f_2(x)$, $f_3(x)$, P_1, P_2, P_3 są dane, a $f_T(x)$, P_T są prognozowane.



Rys. 1

Przykłady

(a) Badano wielkości kosztów x w pojedynczych przedsiębiorstwach j=1, 2,...,N w latach t=1,2,...,n. Na podstawie informacji $\{x_{jt}, j=1,2,...N\}$ o wielkości kosztu w poszczególnych przedsiębiorstwach w roku t ustalono rozkład $f_t(x)$, $x \in P_t$ kosztu w badanej grupie przedsiębiorstw w roku t ($1 \leqslant t \leqslant n$). Interesuje nas odgadnięcie rozkładu kosztów w roku $T > n$.

(b) W n przedsiębiorstwach (t=1,2,...,n) wytwarzających ten sam wyrób obserwowano kształtowanie się częstości wyrobów wadliwych. Wielkości te wiązano z wielkościami technicznego uzbrojenia $y_1, y_2,...,y_n$ w tych przedsiębiorstwach. Dla każdego przedsiębiorstwa z osobna na podstawie sprawozdawczości (np. dziennej) określono rozkład częstości $f_t(x)$, $x \in P_t$ wyrobów wadliwych x. Interesuje nas prognoza rozkładu częstości wyrobów wadliwych w przedsiębiorstwie, w którym techniczne uzbrojenie pracy jest równe y_T.

(c) Dla kolejnych lat t=1,2,...,n skonstruowano modele $f_t(x)$, $x \in P_t$, opisujące kształtowanie się wydatków względem dochodu x w roku t. Chcemy odgadnąć model wydatków względem dochodu w roku $T > n$.

W przedstawionych przykładach z daną wartością y_t zmiennej y wiązano kształt $\{f_t(x), x \in P_t\}$, a zadanie polegało na sporządzeniu prognozy kształtu $\{f_T(x), x \in P_T\}$, związanego z wartością $y=y_T$. W przykładach (a), (c) było $y_t=t$.

Dodajmy, że zagadnienia rozpatrywane w przykładach (a), (b) to zagadnienia prognozowania rozkładów, a w przykładzie (c) - zagadnienie prognozowania modelu.

Dalej omówimy niektóre sposoby prognozowania kształtu $\{f_T(x), x \in P_T\}$ na podstawie kształtów $\{f_t(x), x \in P_t\}$, t=1,...,n. Wielkości prognozowane oznaczać będziemy gwiazdką, np. prognozą funkcji $f_T(x)$ będzie $f_T^*(x)$, a prognozą jej dziedziny P_T będzie P_T^*. Rozróżnimy dwa przypadki

a) gdy kształty $[\{f_t(x), x \in P_t\}$ t=1,...,n,T] są homogeniczne

b) gdy nie zakłada się homogeniczności kształtów.

Niech c_t oznacza wektor parametrów funkcji $f_t(x)$. Powiemy, że kształty $[\{f_t(x), x \in P_t\}, t=1,...,n,T]$ są homogeniczne, jeśli

$$\bigwedge_{t=1,...,n,T} f_t(x) = f(x,c_t)$$

tzn. jeśli funkcje $f_t(x)$ mają tę samą ogólną postać analityczną, a różnię się mogą jedynie wartościami parametrów.

2. Prognozowanie $f_T(x)$, $x \in P_T$ na podstawie prognoz jej parametrów. Kształty homogeniczne

Przyjmijmy, że funkcje $f_t(x)$ oraz ich dziedziny P_t, t=1,...,n są już dane (w praktyce często jest tak, że funkcje te trzeba dopiero wyznaczyć na podstawie materiału statystycznego. Do sprawy tej powrócimy w dalszej części pracy).

Zakładamy, że wszystkie funkcje $f_t(x)$, $t=1,\ldots,n,T$ mają tę samą ogólną postać analityczną, tzn.

$$f_t(x) = f(x,c_t)$$

gdzie c_t jest wektorem parametrów funkcji $f_t(x)$. Zmienna x może być zmienną ciągłą lub dyskretną.

W przypadku kształtów homogenicznych zastosować można następującą procedurę prognozowania kształtu $\{f_T(x),\ x \in P_T\}$:

1^o Niech c_t^r oznacza wartość r-tego parametru funkcji $f_t(x)$, $r=1,\ldots,R$.

Na podstawie informacji o wartościach $[c_t^r,\ t=1,2,\ldots,n]$ sporządzamy prognozę c_T^r wartości r-tego parametru funkcji $f_T(x)$.

2^o Na podstawie informacji o dziedzinach P_T poszczególnych funkcji $f_t(x)$, $t=1,\ldots,n$ sporządzamy prognozę P_T^* dziedziny funkcji $F_T(x)$.

3^o Prognozą kształtu $\{f_T(x),\ x \in P_T\}$ jest

$$\{f_T^*(x),\ x \in P_T^*\}$$

gdzie

$$f_T^*(x) = f(x,c_T^*)\ ,\quad c_T^* = [c_T^{r*}]r=1,\ldots,R$$

Wyjaśnimy poszczególne etapy bardziej szczegółowo.

Prognozowanie parametru c_T^r

Informacja statystyczna wykorzystywana w prognozowaniu wartości c_T^r ma postać szeregu

$$[(c_t^r\ ,\ y_t),\ t = 1,\ldots,n]$$

i będzie to np. dwuwymiarowy szereg czasowy, jeśli y nie jest czasem bądź nawet jednowymiarowy szereg czasowy, gdy zmienna y jest czasem (wówczas $y_t=t$). Ogólnie można więc powiedzieć, że prognoza c_T^r sporządzana jest na podstawie szeregu $[(c_t^r,\ y_t),\ t=1,\ldots,n]$.

Przykładowo, jeśli punkty $[(c_t^r,\ y_t),\ t=1,\ldots,n]$ układają się w otoczeniu pewnej gładkiej krzywej (np. liniowej, wykładniczej, potęgowej itp.), to punkty te aproksymujemy odpowiednio gładką funkcją zmiennej y

$$\hat{c}^r = g_r(y,d_r) \qquad d_r \text{ - wektor parametrów}$$

Za prognozę c_T^r przyjmujemy wartość funkcji $\hat{c}^r$ przy $y=y_T$.

$$c_T^r = g_r(y_T,\ d_r).$$

Jeśli punkty $[(c_t^r,\ y_t),\ t=1,\ldots,n]$ wyraźnie układają się w otoczeniu pewnej funkcji segmentowej, to aproksymujemy je odpowiednią funkcją segmentową zmiennej y, a prognozę c_T^r sporządzamy na podstawie takiej funkcji segmentowej.

Jeśli wreszcie punkty $[(c_t^r, y_t), t=1,\ldots,n]$ układają się „bezładnie", to c_T^r można prognozować za pomocą metody Hellwiga na podstawie aproksymanty ruchomej dopasowanej do punktów $[(c_t^r, y_t), t=1,\ldots,n]$.

Oczywiście stosowane mogą być i inne metody prognozowania c_T^r. Możliwości jest tu wiele, bowiem - powtórzmy - zadanie ustalania prognozy c_T^{r*} to zadanie prognozowania wartości c_T^r na podstawie szeregu $[(c_t^r, y_t), t=1,\ldots,n]$. Teoria prognozowania oferuje tu wiele metod.

Prognozowanie dziedziny P_T

Niekiedy może być tak, że dziedziny poszczególnych funkcji $f_t(x)$, $t=1,\ldots,n$ oraz funkcji $f_T(x)$ są z góry założone i równe P. Wówczas oczywiście $P_T^* = P$. Np. jeśli $f_t(x)$ $t=1,\ldots,n,T$ są funkcjami rozkładu zmiennej losowej X o rozkładzie normalnym, to $P_t = (-\infty, \infty)$ i w konsekwencji $P_T = (-\infty, \infty)$.

Przyjmijmy teraz, że x jest zmienną ciągłą, a dziedziny P_t, $t=1,\ldots,n$, T są przedziałami skończonymi

$$P_t = [a_t, b_t] \qquad a_t > -\infty, \quad b_t < \infty .$$

W tym wypadku sporządzanie prognozy $P_T^* = [a_T^*, b_T^*]$ dziedziny $P_T = [a_T, b_T]$ wymaga sporządzenia prognoz albo krańców przedziału P_T, albo długości przedziału P_T i jednego jego krańca. Gdybyśmy prognozowaną wielkość oznaczyli przez s_T ($s_T = a_T, b_T, b_T - a_T$), to informacją statystyczną wykorzystywaną w prognozowaniu s_T jest szereg $[(s_t, y_t), t=1,\ldots,n]$, gdzie odpowiednio $s_t = a_t, b_t, b_t - a_t$. Prognoza s_T^* wyznaczona byłaby na podstawie szeregu $[(s_t, y_t), t=1,\ldots,n]$ i oczywiście stosowane mogą być różne metody (klasyczne, segmentowe, Hellwiga itd.).

Przykład 1

W pewnej grupie przedsiębiorstw badano rozkłady kosztów jednostkowych w czasie i postanowiono sporządzić prognozę rozkładu kosztów w roku t=11 na podstawie informacji o rozkładach kosztów w latach t=1,2,...,10. Przyjęto, że rozkład kosztów $f_t(x)$ w roku t jest logarytmiczno-normalny

$$f_t(x) = \frac{1}{\sqrt{2\pi}\,\sigma_t\,x} \exp\left\{-\frac{\ln(x) - m_t}{2\sigma_t^2}\right\}, \qquad x \in P_t = (0, \infty)$$

gdzie σ_t, m_t - parametry.

W tabeli 1 podano wartości parametrów σ_t, m_t poszczególnych funkcji $f_t(x)$.

Tabela 1

rok t	Parametry funkcji $f_t(x)$	
	σ_t	m_t
1	0.1	1.3
2	0.2	1.5
3	0.3	1.5
4	0.5	1.7
5	0.6	1.8
6	0.8	1.9
7	1.0	2.0
8	0.8	2.2
9	0.7	2.3
10	0.5	2.3

Prognozowanie parametru σ_T

Po sporządzeniu wykresu przekonujemy się, że ciąg punktów $[(\sigma_t, t), t=1,\ldots,10]$ układa się „segmentowo": pierwszy segment zawiera dane dotyczące t=1,...,6; drugi - dane dotyczące t=7,8,9,10. Przyjmiemy więc, że prognoza σ_T^* sporządzona będzie poprzez ekstrapolację ostatniego segmentu. Dopasowując do danych

t	7	8	9	10
σ_t	1.0	0.8	0.7	0.5

segment liniowy (klasyczną mnk) otrzymujemy, że ów segment ma postać

$$\hat{\sigma}(t) = -0.16t + 2.11 \qquad t = 7,8,9,19 .$$

Tak więc prognozą parametru σ_T dla t=T=11 jest

$$\sigma_T^* = -0.16 \cdot 11 + 2.11 = 0.35 .$$

Prognozowanie parametru m_T

Można przyjąć, że punkty $[(m_t, t), t=1,\ldots,10]$ układają się w otoczeniu pewnej prostej. Dopasowując do tych danych trend liniowy (za pomocą klasycznej mnk) otrzymujemy, że

$$\hat{m}(t) = 0.11576t + 1.2133 \qquad (t=1,\ldots,10) .$$

Ekstrapolując ten trend na czas T=11 mamy

$$m_T^* = 2.487 .$$

Prognoza dziedziny P_T

Rozpatrywane tu rozkłady logarytmiczno-normalne z definicji określone są na przedziałach $P_t = (0,\infty)$. Zatem $P_T^{*} = (0,\infty)$.

Prognoza rozkładu kosztów jednostkowych

Uwzględniając prognozy parametrów funkcji $f_T(x)$ oraz jej dziedziny otrzymujemy następującą prognozę funkcji rozkładu kosztów jednostkowych w roku T=11:

$$f_T^{*}(x) = \frac{1}{\sqrt{2\pi} \cdot 0.35x} \exp\left\{ - \frac{\ln(x) - 2.478}{2 \cdot 0.35^2} \right\} =$$

$$= \frac{1}{0.8771\ x} \exp\left\{ - \frac{\ln(x) - 2.478}{0.245} \right\}, \quad x \in (0,\infty)$$

3. Prognozowanie funkcji $f_T(x)$ na podstawie jej wartości

Jak w poprzednim paragrafie zakładamy, że funkcje $f_t(x)$ oraz ich dziedziny P_t (t=1,...,n) są dane. Zakładamy przy tym, że x jest zmienną ciągłą, a f(x) jest funkcją jednoznaczną przy wszystkich $x \in P_t$. Zakładamy ponadto, że dziedziny P_t są przedziałami skończonymi

$$P_t = [a_t, b_t].$$

Z praktycznego punktu widzenia nie jest to zbyt mocne ograniczenie. W praktyce przedział $(-\infty, \infty)$ zastąpić można bowiem przedziałem $[-a_t, b_t]$, gdzie a_t, b_t są bardzo dużymi liczbami dodatnimi.

Jednocześnie przyjmiemy słabsze niż poprzednio założenie, co do postaci funkcji $f_t(x)$ t=1,...,n,T. Mianowicie nie postulujemy, by kształty były homogeniczne. Przedstawiona dalej procedura prognozowania jest właśnie pomyślana jako procedura prognozowania w przypadku kształtów niehomogenicznych (choć może być też zastosowana, gdy kształty są homogeniczne). Procedura ta sprowadza się do bezpośredniego prognozowania wartości funkcji $f_T(x)$. (Poprzednio, w paragrafie 2, bezpośrednio prognozowano parametry funkcji $f_T(x)$.)

Intencja proponowanej procedury prognozowania jest następująca.

(a) Ponieważ dziedziny poszczególnych funkcji mogą być różne, zatem rzutujemy funkcje $f_t(x)$, t=1,...,n na pewien wcześniej ustalony przedział [a,b].

(b) Na podstawie wartości $f_t'(x)$ rzutów tych funkcji w punkcie $x' \in [a, b]$ ustalamy prognozę dotyczącej punktu x' wartości funkcji $f_T'(x')$, będącej rzutem funkcji $f_T(x)$ na przedział [a,b].

(c) Ustalamy punkt $x \in P_T$, odpowiadający punktowi $x' \in [a,b]$ i przyjmujemy, że prognoza wartości funkcji $f_T(x)$ w tym punkcie $x \in P_T$ jest równa prognozie wartości $f_T'(x')$.

Opiszemy szczegółowo proponowaną procedurę prognozowania kształtu $\{f_T(x), x \in P_T\}$.

1° Prognozujemy przedział P_T

Uwagi na temat prognozowania dziedziny P_T funkcji $f_T(x)$ podano już w poprzednim paragrafie i przeto nie będziemy ich powtarzać.

2° Ustalamy „jednolity" przedział $P = [a,b]$ i wybieramy regułę przekształcenia przedziałów $P_t = [a_t, b_t]$ $t=1,\ldots,n,T$ w przedział P.

Niech x' oznacza punkt z przedziału P, a x punkt z przedziału P_t. Zastosowane mogą być różne reguły przekształcenia przedziałów P_t w przedział P, np. przekształcenie liniowe

$$x' = c_t x - d_t$$

gdzie $c_t = \frac{b - a}{b_t - a_t}$, $d_t = a - a_t c$.

W szczególności może to być przekształcenie liniowe, przekształcające P_t w przedział jednostkowy $P = \{0,1\}$. Wówczas

$$x' = \frac{x - a_t}{b_t - a_t} .$$

Reguła przekształcenia P_t w przedział P jest określona przez wybór pewnej jedno-jednoznacznej funkcji g_t takiej, że

$$x' = g_t(x) , \quad x' \in P , \quad x \in P_t$$

O funkcjach g_t $(t=1,\ldots,n,T)$ zakładamy, że mają tę samą ogólną postać analityczną, a ich parametrami są a_t, b_t:

$$g_t(x) = g(x;\ a_t, b_t) .$$

3° Reprezentujemy kształt $\{f_t(x), x \in P_t\}$ kształtem $r_t(x')$ określonym dla wybranych punktów przedziału P

Operacyjnie obecny etap procedury polega na
- wybraniu ciągu punktów x'_j $(j=1,\ldots,N)$ z przedziału P
- przypisaniu każdemu takiemu x_j wartości

$$r_t(x'_j) = f_t(x_{tj})$$

gdzie punkt

$$x_{tj} = g_t^{-1}(x'_j), \qquad (t=1,\ldots,n)$$

W ten sposób kształt $\{f_t(x), x \in P_t\}$ będzie reprezentowany kształtem $\{r_t(x'_j), j=1,\ldots,N\}$ określonym nad wybranymi punktami przedziału P. Konsekwencją takiej reprezentacji kształtów $\{f_t(x), x \in P_t\}$ $t=1,\ldots,n$ jest możliwość traktowania ciągu wartości

$$r_1(x'_j),\ r_2(x'_j),\ldots,r_n(x'_j)$$

jako odpowiednika ciągu wartości

$$f_1(x_{1j}),\ f_2(x_{2j}),\ldots,f_n(x_{nj})$$

poszczególnych funkcji $f_t(x)$ w punktach $x_{tj} \in P_t$ odpowiadających punktowi $x'_j \in P$.

Jest zrozumiałe, że dla dostatecznie dobrej reprezentacji kształtu $\{f_t(x),\ x \in P_t\}$ kształtem $\{r_t(x'_j),\ j=1,\ldots,N\}$ potrzebny jest dość „gęsty" i „równomierny" wybór punktów x'_j z przedziału P.

4° Prognozujemy wartości $r_t(x'_j)$, $j=1,\ldots,N$

W wyniku poprzedniego etapu dla danego $x'_j \in P$ otrzymujemy ciąg wartości

$$r_1(x'_j),\ r_2(x'_j),\ldots,r_n(x'_j)\ ,$$

które można uznać za wartości porównywalne, gdyż dotyczą tego samego punktu x'_j. Jednocześnie wartość $r_t(x'_j)$ związana jest z wartością y_t zmiennej y, bowiem $r_t(x'_j)$ odpowiada wartości $f_t(x_{tj})$, a ta związana jest z y_t. Nasuwa się zatem następująca sugestia sposobu prognozowania wartości $r_T(x'_j)$, $(1 \leqslant j \leqslant N)$:

- bierzemy pod uwagę szereg $[(r_t(x'_j),\ y_t),\ t=1,\ldots,n]$
- na podstawie tego szeregu sporządzamy prognozę $r^{*}_T(x'_j)$ wartości $r_T(x'_j)$.

Prognoza $r^{*}_T(x'_j)$ może być klasyczna, segmentowa, wg metody Hellwiga itd.

5° Prognozujemy wartości funkcji $f_T(x)$

Niech g_{*T} oznacza funkcję przekształcającą przedział $P^{*}_T = [a^{*}_T,\ b^{*}_T]$ w przedział $P = [a,b]$, przy czym

$$g_{*T}(x) = g(x, a^{*}_T,\ b^{*}_T) \qquad x \in P^{*}_T$$

Prognozą wartości funkcji $f_T(x)$ w punkcie

$$x_{Tj} = g_T^{-1}(x'_j)$$

jest

$$f^{*}_T(x_{Tj}) = r^{*}_T(x'_j), \qquad j=1,\ldots,N\ .$$

Przedstawiona procedura prognozowania pozwala określić prognozy wartości funkcji $f_T(x)$ dla wyżej zdefiniowanych punktów x_{Tj} $(j=1,\ldots,N)$. Jeśli liczba N jest duża, zbiór $\{f^{*}_T(x_{Tj}),\ j=1,\ldots,N\}$ można uznać za praktycznie wystarczającą reprezentację funkcji $f^{*}_T(x)$, $x \in P^{*}_T$.

Gdybyśmy jednak z jakichś względów chcieli otrzymać „kompletną" funkcję $f^{*}_T(x)$ określoną dla wszystkich $x \in P^{*}_T$, to można byłoby to zrobić przynajmniej na dwa sposoby:

a) poprzez interpolację z węzłami interpolacji

$$[f^{*}_T(x_{Tj}),\ x_{Tj}]\ , \qquad [f^{*}_T(x_{Tj+1}),\ x_{T,j+1}]\ ;$$

b) poprzez aproksymację punktów

$$[(f_T^{*}(x_{Tj}),\ x_{Tj}),\ j=1,\ldots,N]$$

funkcją zmiennej x. Tą „nową" funkcję uznajemy za prognozę funkcji $f_T(x)$.

Przykład 2

Bada się zmiany kształtów w czasie (tzn. $y_t = t$). Dla czterech kolejnych lat t=1,2,3,4 ustalono, że

$$f_1(x) = 8e^{0.05x} \qquad \text{przy czym} \qquad P_1 = [1,10]$$
$$f_2(x) = 0.9x + 3 \qquad\qquad\qquad\quad P_2 = [3,15]$$
$$f_3(x) = x + 1 \qquad\qquad\qquad\quad P_3 = [4,18]$$
$$f_4(x) = 3e^{0.1x} \qquad\qquad\qquad\quad P_4 = [6,21]$$

Należy sporządzić prognozę kształtu $f_T(x)$, $x \in P_T$ dla t=5.

1. Prognoza przedziału P_T

Przyjmiemy, że przedział P_T prognozowany jest przez prognozowanie jego krańców: dolnego a_T i górnego b_T. Dolne krańce a_t oraz górne krańce b_t przedziałów P_t (t=1,2,3,4) są następujące

t	1	2	3	4
a_t	1	3	4	6
b_t	10	15	18	21

Ponieważ y=t prognozy a_T^{*}, b_T^{*} wyznaczymy na drodze ekstrapolacji trendów odpowiednio: wartości $\{a_t,\ t=1,2,3,4\}$, wartości $\{b_t,\ t=1,2,3,4\}$. Po sporządzeniu wykresu tych szeregów przekonujemy się, że rozsądnie będzie wziąć trendy liniowe. Stosując klasyczną mnk otrzymujemy

$$\hat{h}_a = 1.6t - 0{,}5\ ; \qquad \hat{h}_b = 3.6t + 7$$

($\hat{h}_a$ - trend wartości a_t, $\hat{h}_b$ - trend wartości b_t).

Ekstrapolując te trendy na czas t=T=5 uzyskujemy

$$a_T^{*} = 1.6 \cdot 5 - 0.5\ ; \qquad b_T^{*} = 3.6 \cdot 5 + 7 = 25\ .$$

Prognozą przedziału P_T jest

$$P_T^{*} = [7.5,\ 25]\ .$$

2. Reguła przekształcenia przedziałów P_t w przedział P

Przyjmiemy, że jednolitym przedziałem jest P = [0,1] oraz, że punktowi $x \in P_t$, w przedziale P odpowiada punkt

$$x' = \frac{x - a_t}{b_t - a_t} \qquad (t=1,\ldots,n,T;\ n=4,\ T=5)$$

3. Reprezentowanie kształtów $f_t(x)$, $x \in P_t$ kształtami $r_t(x')$, $x' \in P$

Wybierzemy następujące punkty x'_j (j=1,...,N) z przedziału P:

j	1	2	3	4	5	6	7	8	9	10	11
x_j	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

, (N=11)

Ponieważ $x' = \frac{x - a_t}{b_t - a_t}$, więc punktowi x'_j z przedziału P odpowiada w przedziale P_t punkt

$$x_{tj} = x'_j (b_t - a_t) + a_t \qquad t=1,...,n .$$

W tabeli 2 podano tak obliczone wartości x_{tj} oraz związane z nimi wartości

$$r_t(x'_j) = f_t(x_{tj}) \qquad t=1,2,3,4 .$$

Tabela 2

j	x'_j	t=1		t=2		t=3		t=4	
		x_{1j}	$f_1(x_{1j})$ $r_1(x'_j)$	x_{2j}	$f_2(x_{2j})$ $r_2(x'_j)$	x_{3j}	$f_3(x_{3j})$ $r_3(x'_j)$	x_{4j}	$f4(x_{4j})$ $r_4(x'_j)$
1	0	1	8.41	3	5.70	4	5	6	5.46
2	0.1	1.9	8.79	4.2	6.78	5.4	6.4	7.5	6.35
3	0.2	2.8	9.20	5.4	7.86	6.8	7.8	9.0	7.38
4	0.3	3.7	9.63	6.6	8.94	8.2	9.2	10.5	8.57
5	0.4	4.6	10.07	7.8	10.02	9.6	10.6	12.0	9.96
6	0.5	5.5	10.53	9.0	11.10	11.0	12.0	13.5	11.57
7	0.6	6.4	11.02	10.2	12.18	12.4	13.4	15.0	13.45
8	0.7	7.3	11.52	11.4	13.26	13.8	14.8	16.5	15.62
9	0.8	8.2	12.05	12.6	14.34	15.2	16.2	18.0	18.15
10	0.9	9.1	12.61	13.8	15.42	16.6	17,6	19.5	21.09
11	1.0	10.0	13.19	15.0	16.50	18.0	19.0	21.0	24.50

Tak więc np. w punktach x_{1j}=1, 1.9, 2.8,...,9.1, 10 kształt $f_1(x)$, $\in P_1$ reprezentowany jest kształtem $r_1(x')$, $x' \in P$ w punktach x'_j=0, 0.1, 0.2,..., 0.9, 1.0. Wartości $f_1(x)$, $r_1(x')$ w odpowiadających sobie punktach x_{1j}, x'_j są równe 8.41, 8.79, 9.20,...,12.61, 13.19.

4. Prognozowanie wartości $r_T(x'_j)$, j=1,...,N

Wartość $r_T(x'_j)$ prognozowana jest na podstawie szeregu $\left[(r_t(x'_j), y_t), t=1,...,n\right]$. W naszym przykładzie zmiany kształtów wiązane są z czasem (y=t), zatem $r_T(x'_j)$ będziemy prognozować na podstawie jednowymiarowego szeregu czasowego

$$\left[r_t(x'_j), \quad t=1,...,n\right] .$$

Ciągi liczb $[r_t(x'_j),\ t=1,2,3,4]$ podano w odpowiednich j-ych wierszach tabeli 2.

Przyjmiemy, że prognozą wartości $r_T(x'_j)$ będzie wartość z ekstrapolacji trendu dopasowanego do szeregu $[r_t(x'_j),\ t=1,2,3,4]$ i przy tym będą to trendy

- hiperboliczne dla j=1,2,3
- liniowe dla j=4,5,...,11.

Po zastosowaniu klasycznej mnk otrzymano następujące trendy $\hat{r}_j$ wartości $[r_t(x'_j),\ t=1,2,3,4]$:

$$\hat{r}_1 = \frac{4.14}{t} + 3.89, \quad \hat{r}_2 = \frac{3,39}{t} + 5.31, \quad \hat{r}_3 = \frac{2.31}{t} + 6.86$$

$$\hat{r}_4 = -0.262t + 9.82, \quad \hat{r}_5 = 0.025t + 10.1, \quad \hat{r}_6 = 0.402t + 10.3$$

$$\hat{r}_7 = 0.851t + 10.39, \quad \hat{r}_8 = 1.322t + 10.34, \quad \hat{r}_9 = 2.016t + 10.145$$

$$\hat{r}_{10} = 2.762t + 9.775, \quad \hat{r}_{11} = 3.643t + 9.19 .$$

Prognozy wartości $r_T(x'_j)$ j=1,...,11 otrzymane z ekstrapolacji odpowiednich trendów na czas T=5 podano w tabeli 3.

Tabela 3

j	1	2	3	4	5	6	7	8	9	10	11
x'_j	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$r_T(x'_j)$	4.72	5.99	6.92	8.51	10.23	12.31	14.65	16.95	20.23	23.59	27.41

5. Prognoza kształtu $\{f_T(x),\ x \in P_T\}$

Ponieważ

$$x' = \frac{x - a_t}{b_t - a_t}, \qquad P = [0,2],$$

więc punktowi x'_j z jednolitego przedziału P = [0,1] odpowiada punkt

$$x_{Tj} = x'_j(b^*_T - a^*_T) + a^*_T$$

z przedziału $P_T = [a^*_T, b^*_T]$, będącego prognozą przedziału P_T. Przypomnijmy, że otrzymaliśmy $P^*_T = [7.5, 25]$, zatem

$$x_{Tj} = x'_j(25 - 7.5) + 7.5$$

Prognozą wartości $f_T(x)$ w takim punkcie x_{Tj} jest

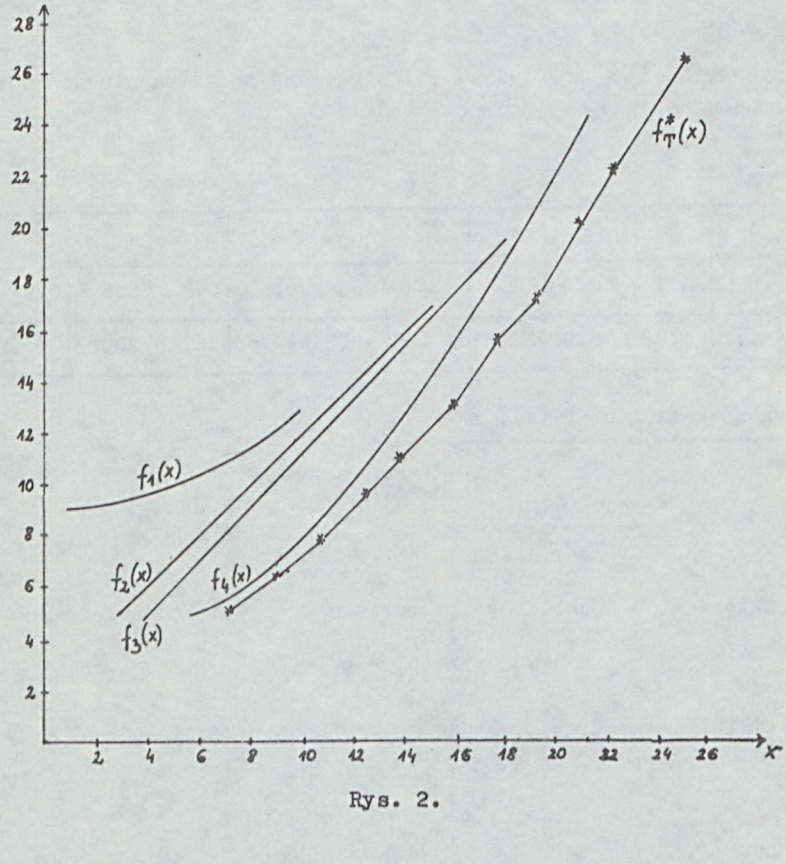
$$f^*_T(x_{Tj}) = r^*_T(x'_j) \qquad j=1,...,N .$$

Wartości x_{Tj} oraz prognozy $f_T^{\#}(x_{Tj})$ podano w tabeli 4.

Tabela 4

x'_j	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
x_{Tj}	7.5	9.25	11.0	12.75	14.5	16.25	18.0	19.75	21.5	23.25	25
$f_T(x_{Tj})$	4.72	5.99	6.92	8.51	10.23	12.31	14.65	16.95	20.23	23.59	27.41

Prognozy te umieszczono też na rys. 2. Dla „uciąglenia" prognozy kształtu $\{f_T^{\#}(x),\ x \in P_T^{\#}\}$ sąsiednie wartości $f_T^{\#}(x_{Tj})$ połączono ze sobą, co odpowiada interpolacji liniowej. Na rys. 2 podano też kształty $f_t(x)$, $x \in P_t$, $t=1,2,3,4$.



Rys. 2.

4. Wyznaczanie funkcji $f_t(x)$, $t=1,\ldots,n$ na podstawie materiału statystycznego

Do tej pory zakładaliśmy, że kształty $\{f_t(x),\ x \in P_t\}$, $t=1,\ldots,n$ są już dane. Obecnie zajmiemy się zagadnieniem wyznaczania funkcji $f_t(x)$,

$t=1,\ldots,n$ na podstawie materiału statystycznego. Tak jak w poprzednich paragrafach założymy, że $f_t(x)$ są funkcjami jednoznacznymi określonymi dla $x \in P_t$. Omówimy trzy „typowe" przypadki konstrukcji funkcji $f_t(x)$.

1. $f_t(x)$ jest funkcją rozkładu zmiennej losowej X

W tym przypadku trzeba ustalić ogólną postać analityczną funkcji $f_t(x)$ i stosowną metodę - np. metodę momentów - wyznaczyć parametry funkcji $f_t(x)$ na podstawie związanych z wartością y_t realizacji zmiennej losowej X.

Przykładowo, jeśli zakładamy, że $f_t(x)$ są funkcjami gęstości rozkładu normalnego, tzn. jeśli zakładamy, że

$$f_t(x) = \frac{1}{\sigma_t \sqrt{2\pi}} \exp\left\{ -\frac{(x-m_t)^2}{2\sigma_t^2} \right\}, \quad x \in P_t = (-\infty, \infty) ,$$

gdzie m_t, σ_t - parametry, to na podstawie metody momentów przyjmiemy, iż

$$m_t = \bar{x}_t , \qquad \sigma_t = s_t$$

gdzie $\bar{x}_t$ - średnia arytmetyczna, s_t - odchylenie standardowe z próby.

2. Aproksymacja empirycznych odpowiedników wartości funkcji $f_t(x)$

Przyjmujemy, że dla danej wartości y_t, z którą związany jest kształt $\{f_t(x),\ x \in P_t\}$ dany jest

- ciąg $[x_u^t\ ;\ u=1,\ldots,u_t]$ wartości zmiennej x
- oraz ciąg $[z_u^t\ ;\ u=1,\ldots,u_t]$ empirycznych odpowiedników wartości funkcji $f_t(x)$ w punktach $x=x_u^t$ $(u=1,\ldots,u_t)$.

Na przykład, jeśli $f_t(x)$ jest modelem ekonometrycznym, to z_u^t będzie wartością zmiennej objaśnianej, a x_u^t - wartością zmiennej objaśniającej.

Jeśli $f_t(x)$ jest charakterystyką rozkładu zmiennej losowej X, to z_u^t będzie empiryczną częstością przyporządkowaną wartości x_u^t. W tym ostatnim przypadku częstości mogą być skumulowane ($f_t(x)$ będzie wówczas dystrybuantą) lub nieskumulowane ($f_t(x)$ będzie funkcją gęstości lub funkcją rozkładu prawdopodobieństwa).

Jeśli $f_t(x)$ jest rozkładem liczebności, to z_u^t będzie liczebnością przypadków przyporządkowanych wartości x_u^t zmiennej x. Liczebności te mogą być skumulowane bądź nieskumulowane.

Jest zrozumiałe, że jeśli zmienna x (bądź zmienna losowa X) jest ciągła, to należy rozpatrywać

- albo liczebności (częstości) skumulowane
- albo liczebności (częstości) nieskumulowane dotyczące pewnego przedziału wartości zmiennej x (zmiennej X), którego reprezentantem jest x_u^t.

Obecnie rozważane podejście polega na tym, że za funkcję $f_t(x)$ przyjmujemy funkcję aproksymującą punkty

$$[(z_u^t, x_u^t),\ u=1,\ldots,u_t] .$$

3. Określanie $f_t(x)$ bez aproksymacji jej empirycznych odpowiedników

Funkcję $f_t(x)$ można również określić na podstawie wartości z_u^t $(u=1,\ldots ...u_t)$ jej empirycznych odpowiedników w punktach x_u^t, bez konieczności aproksymacji. Możemy mianowicie przyjąć, że

a) w przypadku zmiennej x (bądź zmiennej losowej X) dyskretnej

$$f_t(x) = z_u^t \qquad \text{dla } x=x_u^t \qquad (u=1_t,\ldots,u_t)$$

b) w przypadku zmiennej x (bądź zmiennej X) ciągłej

$$\begin{cases} \text{dla } x=x_u^t \quad \text{jest} \quad f_t(x) = z_u^t \\ \text{dla } x \in \langle x_u^t , x_{u+1}^t \rangle \text{ wartość } f_t(x) \text{ jest interpolacją wartości } z_u^t, z_{u+1}^t \end{cases}$$

Takie rozwiązanie wydaje się korzystniejsze od poprzedniego przynajmniej z tego powodu, że dla $x=x_u^t$ wartość $f_t(x)$ jest równa z_u^t (poprzednio $f_t(x)$ była tylko aproksymacją wartości z_u^t).