

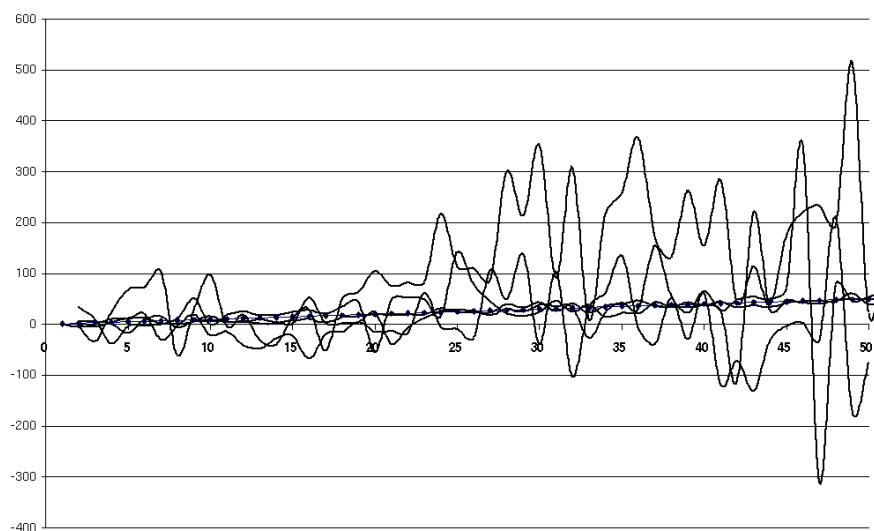
Determinizm osobliwością probabilisty

Andrzej DĄBROWSKI, Wrocław

Mój odczyt jest swoistą zemstą na deterministach, bo dla deterministy probabilista jest osobą osobliwą, jeśli by nie powiedzieć podejrzaną. W związku z tym wykonam operację odwrotną i opowiem, jak wygląda determinizm w oczach probabilisty.

Determinizm kojarzy mi się z możliwością jednoznacznego przewidzenia, co się stanie w przyszłości, gdy wiemy co się stało do danej chwili. A więc musi się on wiązać z pojęciem czasu i procesami, rozwijającymi się w czasie. Będę więc mówić dziś o procesach stochastycznych.

Proces stochastyczny zmiennej t o wartościach zespolonych wyobrażam sobie jako grubą funkcję



składającą się z mnóstwa funkcji zespolonych zmiennej t .

W chwili t probabilista nie wie, na której funkcji się znajduje, mimo iż zna jej aktualną wartość. Na początku świata za pomocą mechanizmu losowego wylosowano funkcję $x(t, \omega)$, której wartości możemy obserwować. Indeks funkcji, oznaczony literą ω identyfikuje, którą funkcję wylosowano. Taka funkcja nazywa się *trajektorią procesu*. Jeżeli zatrzymamy się w jednym momencie i będziemy obserwować, jakie możliwe wartości przyjmą wszystkie funkcje, które występują w procesie stochastycznym, to otrzymamy zmienną losową o określonym rozkładzie. Charakterystycznymi wielkościami, opisującymi typowe zachowanie się zmiennej losowej w punkcie t są wartość oczekiwana

$$\mu(t) = Ex(t)$$

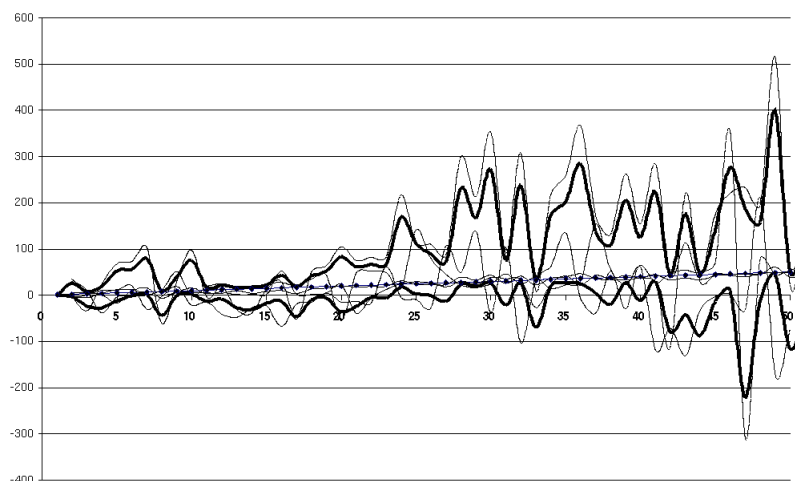
i jej odchylenie standardowe

$$\sigma(t) = \sqrt{E((x(t) - \mu(t))^2)}.$$

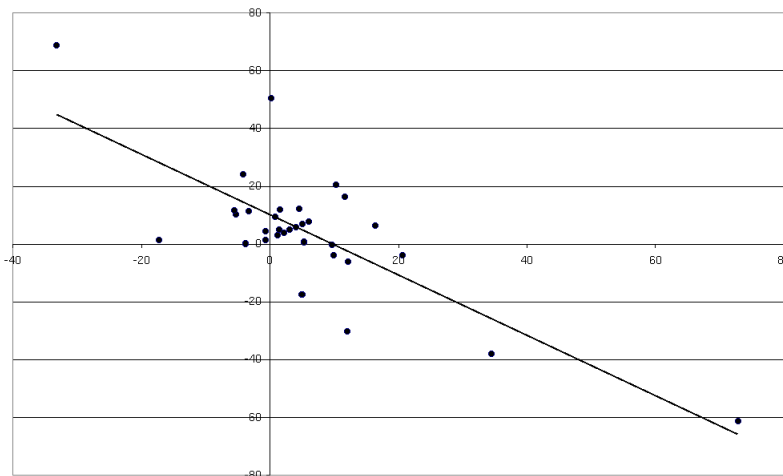
Będziemy zajmować się tylko takimi procesami, dla których istnieje w każdym punkcie odchylenie standardowe. Wtedy także istnieje wartość oczekiwana.

Funkcja $\mu(t)$, zwana wartością oczekiwaną procesu, jest osią, wokół której gromadzą się wszystkie jego trajektorie. Odchylenie standardowe można interpretować jako typowy rozrzut zmiennej wokół wartości oczekiwanej.

Pasek o grubości $\sigma(t)$ wokół $\mu(t)$ zawiera trajektorie najczęściej występujące.



Aby w pełni opisać proces stochastyczny, trzeba ustalić, jak wiążą się z sobą wartości procesu w chwilach t i $t + h$. Jeżeli spojrzymy na funkcję oczyma deterministy, to wie on, że w chwilach t i $t + h$ są wyznaczone jej wartości i to jednoznacznie. Natomiast probabilista jest nieco skonfundowany, bo nie wie, jakich wartości należy oczekiwać w chwili t , a jakich w chwili $t + h$. Aby sobie to wyobrazić, odłóżmy na jednej osi wszystkie możliwe wartości procesu w chwili t a na drugiej – w chwili $t + h$. Punktów o współrzędnych $(x(t, \omega), x(t + h, \omega))$ może być bardzo dużo i pojawiają się one na płaszczyźnie według danego rozkładu prawdopodobieństwa. Jeżeli punkty te leżą w pobliżu jakiejś prostej, to mówimy, że jest silny związek między tym, co się dzieje w chwili t i w chwili $t + h$.



Jeżeli te punkty wytworzą chmurę, z jednorodnie pojawiającymi się punktami w każdym kierunku, to mówimy, że zależność jest słaba: cokolwiek nie działałoby się w chwili t , to praktycznie wszystko niezależnie może się dzieć w chwili $t + h$.

Miarą tej zależności jest *funkcja kowariancyjna* procesu w chwilach t i $t + h$:

$$\gamma(t, t + h) = E((x(t) - \mu(t))(x(t + h) - \mu(t + h))).$$

Zbiór zmiennych losowych można wyobrazić sobie jako przestrzeń liniową z iloczynem skalarnym

$$\langle U, V \rangle = E((U - EU)(\overline{V - EV})).$$

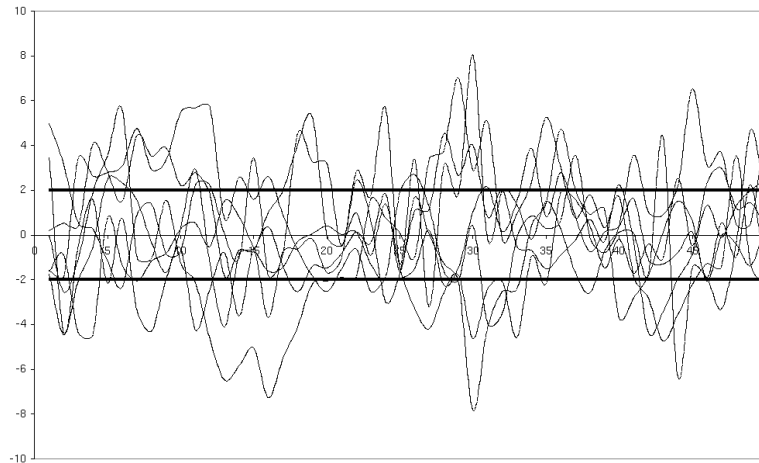
Odchylenie standardowe jest w tej przestrzeni długością (normą) zmiennej losowej:

$$\sigma(U) = \|U\|.$$

Jest to więc przestrzeń Hilberta.

Świat zwykłych funkcji jest skomplikowany, świat zaś grubych funkcji, czyli procesów stochastycznych, jest jeszcze bardziej skomplikowany. Zajmiemy się więc pewnymi szczególnymi procesami, które być może z punktu widzenia deterministy są mało ciekawe.

Ograniczymy się do takich procesów stochastycznych, w których oś procesu, reprezentowana przez wartość oczekiwaną, jest funkcją stałą. Bez straty ogólności będziemy zakładać, że nasz proces ma wartość oczekiwaną 0. Zakładamy ponadto, że tasiemka typowych trajektorii procesu mieści się w prostoliniowym pasku, równoległym do wartości średniej, a więc zakładamy, iż odchylenie standardowe jest stałe w czasie ($\sigma(t) = \sigma$).

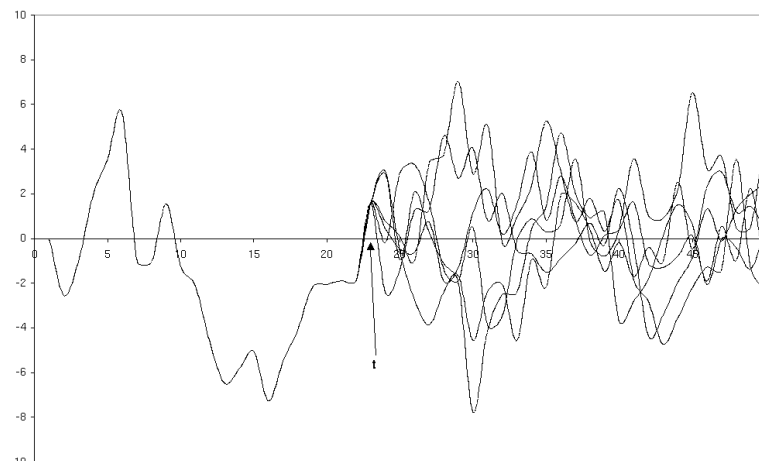


Założymy ponadto, że moment w którym zaczynamy obserwować proces nie ma wpływu na jego przebieg. Oznacza to, że dwa raporty o procesie, zaczynające się w różnych momentach czasowych byłyby, z punktu widzenia własności probabilistycznych, nieodróżnialne. Warunek ten wyrazić można żądając, aby funkcja kowariancyjna procesu $\gamma(t, t + h)$ zależała jedynie od h :

$$\gamma(t, t + h) \stackrel{df}{=} \gamma(h).$$

Oznacza to pewien luksus dla naukowca, który może się włączyć w obserwowanie świata w dowolnym momencie, bo nie zmienia to opisu tego świata. Taki proces nazywa się *stacjonarnym*. Założymy ponadto, co nie zmienia istoty faktów, które w dalszym ciągu przedstawimy, że proces jest notowany tylko w całkowitych momentach czasowych.

Jak już powiedzieliśmy na początku, pojęcie determinizmu wiąże się z możliwością przewidzenia, co się będzie działo w przyszłości. Trajektoria procesu stochastycznego w chwili t na ogół nie wyznacza jednoznacznie wartości w chwili $t + h$, bo może ona przebiegać od chwili t na wiele sposobów.



Podanie wartości jaką będzie miał proces w chwili $t + h$, gdy znamy jego przebieg do chwili t , nazywa się *prognozą* $\hat{x}(t, h)$ w chwili t z horyzontem h . Determinizm oznacza dla nas sytuację, kiedy umiemy podać bezbłędną prognozę, to znaczy, gdy zawsze $\hat{x}(t, h) = x(t + h)$. Dla probabilisty taka sytuacja jest osobliwa i takie procesy nazywa się *całkowicie deterministycznymi* albo *singularnymi*.

Jest jeszcze druga skrajność, gdy najlepsza prognoza z horyzontem h zmierza do średniej wartości procesu dla $h \rightarrow \infty$, czyli $\hat{x}(t, h) \rightarrow 0$ dla $h \rightarrow \infty$. Jedyną rozsądną prognozą z bardzo dalekim horyzontem dla takich procesów jest wartość typowa procesu. Procesy, w których takie prognozy są optymalne, nazywane są *regularnymi*.

Co to znaczy zaproponować dobrą prognozę? Prognoza jest dobra, jeśli jest bliska rzeczywistości. Innymi słowy mówiąc, musi być ona oparta na historii procesu do chwili t i odległość między prognozą a rzeczywistą wartością procesu

$$\|\hat{x}(t, h) - x(t + h)\|$$

musi być jak najmniejsza.

Co to jest historia procesu do chwili t ? Z informacji dostępnych do tej chwili, a więc z rodziny zmiennych losowych $\{x(u) : u \leq t\}$ (są to wszystkie przekroje grubej funkcji do chwili t), wytwarzamy półprodukty, z których będziemy chcieli przyrządzić recepty na prognozę. Półprodukty te powstają przez kombinacje liniowe na tych zmiennych losowych wraz z możliwymi przejściami granicznymi. Najlepsza prognoza $\hat{x}(t, h)$ należy do podprzestrzeni liniowej

$$H_t = \overline{\text{lin}\{x(u) : u \leq t\}}$$

i spełnia warunek

$$\|\hat{x}(t, h) - x(t + h)\| = \min \{\|U - x(t + h)\| : U \in H_t\}.$$

Z geometrii wiemy, jak znaleźć rozwiązanie tego zagadnienia: prognoza jest rzutem prostopadłym wartości procesu w chwili $t + h$ na liniową podprzestrzeń H_t historii procesu do chwili t . Odległość $x(t + h)$ od rzutu, równa $\|\hat{x}(t, h) - x(t + h)\|$ jest niczym innym tylko błędem prognozy, a dokładnie mówiąc, jego odchyleniem standardowym.

Skoro się tak zagłęбилиśmy w historię, to zagłębmy się jeszcze w prehistorię. Wiedza na początku świata jest częścią wspólną wszystkich historii, czyli jest równa

$$H_{-\infty} = \bigcap_t H_t.$$

Ta podprzestrzeń liniowa jest wspólnym jądrem tego, co wiemy o naszym świecie. Jest ono niepuste – w początku świata musi się znajdować co najmniej zmienna losowa, przyjmująca z prawdopodobieństwem 1 wartość 0 (deterministyczna!!). Ale co jeszcze?

Inna przestrzeń liniowa opisuje całą wiedzę, jaką potrafimy zgromadzić o naszym procesie stochastycznym:

$$H_{\infty} = \bigcup_t H_t.$$

Za pomocą tych przed chwilą zdefiniowanych przestrzeni możemy scharakteryzować procesy singularne i regularne.

Twierdzenie

Proces stacjonarny jest całkowicie deterministyczny (osobliwy, singularny) wtedy i tylko wtedy, gdy

$$H_{-\infty} = H_{\infty},$$

a regularny wtedy i tylko wtedy, gdy

$$H_{-\infty} = \{0\}.$$

Pierwsza teza oznacza, że proces jest całkowicie deterministyczny, gdy na początku wszystko jest wiadome. Druga teza mówi, że proces jest regularny gdy na początku nasza wiedza jest minimalna.

W pewnym momencie stwierdziliśmy nie dość przekonująco, że singularność i regularność są to dwie skrajności. Teraz dopiero widać, że rzeczywiście to są dwie skrajności. Jak to często bywa, okazuje się, że świat jest zbudowany wyłącznie ze skrajności. Zachodzi bowiem twierdzenie

Twierdzenie Wolda

1. *Każdy proces stacjonarny $x(t)$ jest sumą procesu singularnego $s(t)$ i regularnego $r(t)$:*

$$x(t) = s(t) + r(t)$$

i ta suma jest wyznaczona w sposób jednoznaczny (z dokładnością do zdarzeń o prawdopodobieństwie 0).

2. *Procesy $s(t)$ i $r(t)$ są ortogonalne.*

- 3.

$$\bigwedge_t (H_t^s \subset H_t^x, H_t^r \subset H_t^x),$$

gdzie H_t^s, H_t^r, H_t^x są historiami procesów $s(t)$, $r(t)$ i $x(t)$.

Punkt 2. twierdzenia Wolda oznacza w języku probabilistycznym, że oba procesy są z sobą nieskorelowane, czyli że te dwie składowe opisują zupełnie inne kawałki świata. Punkt 3. mówi, że informacje z części singularnej i regularnej procesu x są mu podporządkowane w tym sensie, że w każdym momencie wiedza o przeszłości procesów s i r nie jest bogatsza od wiedzy o przeszłości procesu x .

Z twierdzenia Wolda wynika wprost, że aby nauczyć się prognozować procesy stacjonarne wystarczy umieć:

- wydzielić jego część singularną (lub regularną, gdyż druga część jest różnicą procesu i wydzielonej już części),
- prognozować procesy singularne i regularne.

Końcowa prognoza jest sumą prognoz części singularnej i regularnej.

Aby nie poruszać się w próżni zobaczmy, jak mogą wyglądać procesy singularne i regularne.

Przykłady

1. $x(t) = \varepsilon_t$ gdzie ε_t jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie o średniej 0 i odchyleniu standardowym σ .

Takimi procesami zajmuje się klasyczny rachunek prawdopodobieństwa i statystyka. Wartości tego procesu w kolejnych momentach są niezależne i mają taki sam rozkład. Najlepsza prognoza jest zawsze równa wartości oczekiwanej 0, a więc jest to proces regularny.

2. Proces o wartościach zespolonych

$$x(t) = A(\omega)e^{it\lambda_0}$$

gdzie A jest zmienną losową o wartości oczekiwanej 0 i odchyleniu standardowym σ . Część rzeczywista tego procesu jest cosinusoidą, a część urojona sinusoidą o okresie $\frac{2\pi}{\lambda_0}$ i losowej amplitudzie A .

Jest to proces singularny. Gdy wiemy, co się dzieje do chwili t , co więcej, wystarczy, że wiemy, co się dzieje w chwili t , to wiemy, jaka jest amplituda. Gdy to już wiemy, to potrafimy podać bezbłędną prognozę

$$\widehat{x(t, h)} = A(\omega)e^{i(t+h)\lambda_0}.$$

Przykład ten pokazuje, że determinizm wcale nie musi oznaczać braku losowości.

Proces regularny ma bardzo prostą strukturę. Daje się zapisać jako kombinacja liniowa wartości procesu występującego w przykładzie 1.

Twierdzenie

Niech $r(t)$ będzie procesem regularnym. Jest on kombinacją liniową wartości procesu białego szumu:

$$(1) \quad r(t) = \sum_{s \leq t} c(t-s)\varepsilon_s.$$

Prognoza takiego procesu wyraża się wzorem

$$r(\widehat{t, h}) = \sum_{s \leq t} c(t+h-s)\varepsilon_s.$$

Proces stacjonarny można opisać jeszcze inaczej. Własności probabilistyczne procesu wyraża funkcja kowariancyjna $\gamma(h)$, która jest nieujemnie określona, to znaczy spełnia nierówność

$$\sum_{i,j} a_i \gamma(h_i - h_j) \bar{a}_j \geq 0$$

dla dowolnych h_i oraz zespolonych a_i . Ta własność funkcji kowariancyjnej pozwala nam skorzystać z twierdzenia Bochnera, a w zasadzie z jego szczególnego przypadku, zwanego lematem Herglotza. Lemat opisuje, jaka jest postać wszystkich ciągów nieujemnie określonych.

Lemat (Herglotz)

Ciąg nieujemnie określony można przedstawić w postaci

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} \mu(d\lambda),$$

gdzie μ jest pewną miarą na odcinku $(-\pi, \pi)$.

Gdy miara μ ma gęstość $f(\lambda)$ względem miary Lebesgue'a, to

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda) d\lambda.$$

Jak należy interpretować tezę tego lematu? Związek między wartościami procesu w chwilach t oraz $t+h$ opisuje funkcja $\gamma(h)$. Funkcję tę z kolei charakteryzuje miara μ , która mówi z jaką wagą należy wziąć częstość λ w funkcji okresowej $e^{ih\lambda}$ w całce wyznaczającej γ . Lemat ten opisuje w matematycznym języku rozszczepienie światła (niekoniecznie białego) na składowe. Dlatego rozkład w lemacie Herglotza nazywa się *rozkładem spektralnym z miarą spektralną μ* i z *gęstością spektralną $f(\lambda)$* (o ile istnieje).

Obchodzimy właśnie jubileusz Szkół Matematyki Poglądowej. Dla mnie ten odczyt też jest jubileuszowy. Debiutowałem na słynnej, probabilistycznej Szkole nr 2 i mój pierwszy odczyt nazywał się *Kolorowa probabilistyka*, gdzie mówiłem o tym jak kolor, wyznaczony przez rozkład spektralny, wpływa na własności procesu: co to są procesy czerwone, niebieskie, zimne i ciepłe. Przypadek zrządził, że o podobnych sprawach mówię 30 Szkół później.

Wracając do naszego zagadnienia: własności procesu można wyrazić przez własności miary spektralnej, a jeżeli ta miara ma gęstość, to przez własności gęstości spektralnej. Mówiąc w tym języku, ciąg niezależnych zmiennych losowych ma stałe spektrum, a więc może być porównany do światła białego. Dlatego jest nazywany białym szumem. Każdy inny proces jest bardziej kolorowy. Proces singularny, opisany w przykładzie 2, jest procesem o stabilnym kolorze, odpowiadającym częstości λ_0 .

Proces regularny opisany równaniem (1) ma gęstość spektralną

$$f(\lambda) = \left| \sum_{t \geq 0} c(t) e^{-it\lambda} \right|^2.$$

W języku rozkładu spektralnego można też scharakteryzować procesy deterministyczne. Jak wiadomo z teorii miary, każdą miarę μ można rozbić na część, która ma gęstość względem miary Lebesgue'a i na część osobliwą,

skupioną na zbiorze, którego miara Lebesgue'a jest równa 0. Oznacza to, że dla każdego mierzalnego zbioru B

$$\mu(B) = \int_B f(\lambda) d\lambda + \eta(B \cap \Delta_0),$$

gdzie f jest gęstością miary μ względem miary Lebesgue'a, Δ_0 zbiorem miary Lebesgue'a 0, na którym skupiona jest część singularna miary μ .

Przy tych oznaczeniach, prawdziwe jest

Twierdzenie

Stacjonarny proces $x(t)$ o mierze spektralnej μ jest deterministyczny wtedy i tylko wtedy, gdy zachodzi jeden z warunków:

$$f(\lambda) = 0 \text{ dla } \lambda \notin \Delta_0, \quad \text{albo} \quad \int_{-\pi}^{\pi} \ln f(\lambda) d\lambda = -\infty.$$

Proces $x(t)$ jest regularny wtedy i tylko wtedy, gdy $f(\lambda) = 0$ dla $\lambda \in \Delta_0$ i

$$(2) \quad \int_{-\pi}^{\pi} \ln f(\lambda) d\lambda > -\infty.$$

Osobliwość procesów deterministycznych w języku miar spektralnych oznacza, że albo ta miara jest skoncentrowana na zbiorze zerowym względem miary Lebesgue'a (co praktycznie jest równoważne determinizmowi, rozumianemu jako *brak losowości*), albo ta losowość jest bardzo nieznacząca, bo miara spektralna ma tak małe wartości, że całka z jej logarytmu jest równa $-\infty$.

Nierówność (2) związana z procesami regularnymi tylko pozornie wygląda na warunek techniczny. Kołmogorow udowodnił bardzo ciekawe twierdzenie

Twierdzenie

Jeżeli proces $x(t)$ jest regularny z gęstością spektralną f , to

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln(2\pi f(\lambda)) d\lambda = \ln(\sigma^2),$$

gdzie σ^2 jest wariancją błędu prognozy z horyzontem 1.

Z twierdzenia Kołmogorowa wynika, że procesy regularne o małych wartościach gęstości spektralnej można dobrze prognozować. Małe wartości gęstości spektralnej oznaczają, że wartości takiego procesu mają małą zmienność. Stąd wynika, że im bardziej proces jest deterministyczny, tym łatwiej go prognozować. Warunkiem idealnej prognozy może być tylko determinizm.

Może nas interesować jeszcze jeden aspekt determinizmu, związany z *interpolacją*. Obserwujemy proces x do chwili t , ale nie znamy jego wartości w chwilach leżących w skończonym zbiorze $T \subset (-\infty, t]$. Każdy obserwator chciałby te dziury załatać jakimiś sensownymi wartościami. Problem znalezienia dobrej interpolacji jest bardzo podobny do zagadnienia dobrej prognozy. Trzeba załatać dziury w obserwacjach tak, aby błąd, wynikający z interpolacji był jak najmniejszy.

Jeżeli x jest procesem deterministycznym, to najlepsza interpolacja jest bezbłędna. Rozwiązanie jest oczywiste. Na lewo od pierwszej dziury znam całą przeszłość. Proces jest deterministyczny, więc załatał tę dziurę bezbłędnie poprzez prognozę wartości w tym punkcie. Teraz na lewo od drugiej dziury już znam (bzebłędnie!) całą przeszłość, więc przez prognozę wypełnię drugą lukę bezbłędnie. Iterując to postępowanie oszacuję bezbłędnie wszystkie brakujące wartości. Można powiedzieć, że dla procesu deterministycznego zachodzi własność *determinizmu interpolacyjnego*: umiemy załatać bezbłędnie każdy skończony zestaw dziur.

Czy ten rezultat da się przenieść na inne procesy? Nie należy się spodziewać, że tak się da zawsze zrobić. Gęstość spektralna procesu regularnego, dla którego

zachodzi determinizm interpolacyjny musi być w pewnym sensie podobna do gęstości procesu deterministycznego, czyli nie może mieć zbyt dużych wartości. Warunek ten precyzyjnie opisuje następujące

Twierdzenie

Jeżeli proces $x(t)$ jest regularny z gęstością spektralną f , to jest on interpolacyjnie deterministyczny wtedy i tylko wtedy, gdy

$$(3) \quad \int_{-\pi}^{\pi} \frac{|w(\lambda)|^2}{f(\lambda)} d\lambda = \infty,$$

gdzie $w(\lambda) = \sum_k a_k e^{i\lambda k}$ jest dowolną skończoną kombinacją zespolonych jednomianów trygonometrycznych $e^{i\lambda k}$.

Ten warunek jest bardzo mocny: dla każdego wielomianu całka (3) jest osobliwa (znów osobliwość, związana z determinizmem). Procesy regularne interpolacyjnie deterministyczne mają wartości mniejsze niż dowolny wielomian trygonometryczny. Mimo tak trudnego warunku, procesy regularne, dla których determinizm interpolacyjny występuje, istnieją.

Gdy tak nie jest, to istnieje taki wielomian trygonometryczny, że całka (3) jest skończona. Weźmy rodzinę wszystkich takich wielomianów. Można zauważyć, że wszystkie te wielomiany mają wspólny dzielnik i ten wielomian będziemy nazywać wielomianem minimalnym. Ten wielomian jest charakterystyczny dla procesu, bo jest on niejako dzieckiem gęstości spektralnej f . Jest on postaci

$$w_0(\lambda) = \prod_{j=1}^r (e^{i\lambda} - e^{i\lambda_j})^{k_j},$$

gdzie λ_j jest zerem krotności n_j gęstości spektralnej,

$$k_j = \left\lfloor \frac{n_j + 1}{2} \right\rfloor.$$

Mówimy, że λ_0 jest zerem o krotności n gęstości spektralnej f , gdy zachodzi warunek

$$\int_{-\pi}^{\pi} \frac{|\lambda - \lambda_0|^{n-1}}{f(\lambda)} d\lambda = \infty, \quad \int_{-\pi}^{\pi} \frac{|\lambda - \lambda_0|^n}{f(\lambda)} d\lambda < \infty.$$

Wielomian ten pełni szczególną rolę w sytuacji, gdy chcemy wypełnić lukę w obserwacjach w momentach $\{t_0, t_0 + 1, \dots, t_0 + p - 1\}$. Taka dziura da się załatać bezbłędnie wtedy i tylko wtedy, gdy

$$p < \sum_{j=1}^r k_j.$$

Oznacza to, że gdy luka w obserwacjach nie jest zbyt duża, to możemy ją bezbłędnie wypełnić wartościami procesu. Wielkość tej luki nie może przekraczać sumy krotności wielomianu minimalnego, opisującego konfigurację zer gęstości spektralnej.

Literatura

Azencott, R., Dacunha-Castelle, D., *Séries d'observations irrégulières: modélisation et prévision*. Masson, 1984.

Rozanov, Y.A., *Stacjonarnyje sluczajnyje processy*. Fizmatgiz, 1963
tłumaczenia na angielski:

Azencott, R., Dacunha-Castelle, D., *Series of Irregular Observations: Forecasting and Model Building*. Springer, 1986

Rozanov, Y.A. (1967), *Stationary Random Processes*. Holden Day, 1967.