Grzegorz SMOŁALSKI
Politechnika Wrocławska
Wydział Podstawowych Problemów Techniki
Katedra Inżynierii Biomedycznej

ROZKŁADY PRAWDOPODOBIEŃSTWA O OGRANICZONYM NOŚNIKU – PRZYKŁADY ZASTOSOWAŃ W METROLOGII I WYBRANE NASTĘPSTWA TEORETYCZNE

Podano przykładowe powody, dla których stosowanie rozkładów o ograniczonym nośniku jest uzasadnione. Zdefiniowano rozpiętość nośnika. Omówiono wpływ ograniczoności nośnika wielkości wejściowej na rozkład wyjściowy w odwzorowaniach funkcyjnych i w analizach bayesowskich, a także wpływ na kurtozę rozkładu i na szerokość przedziału pokrycia w analizie niepewności.

Słowa kluczowe: funkcja wiarygodności; rozkład *a posteriori*; sprzeczność informacji.

APPLICATION OF THE LIMITTED SUPPORT PROBABILITY DISTRIBUTIONS IN MEASUREMENT SCIENCE – EXAMPLES AND SOME THEORETICAL CONSEQUENCES

Exemplary reasons are given for applying of the limited support distributions in measurements. The range of the support is defined. The influence of the support limitation of the input quantity on the support of the output quantity is discussed both for projections in the form of a function and in Bayessian analyses. Then, also the influence on kurtosis and on the width of the coverage interval in the uncertainty analysis is examined.

Keywords: likelihood function; *a posteriori* distribution; conflicting information items.

1. POWODY I PRZYKŁADY OGRANICZANIA NOŚNIKA; ROZPIĘTOŚĆ NOŚNIKA

Miarą naszego przekonania co do tego, że menzurandowi przynależą wartości z różnych przedziałów jest zwykle rozkład prawdopodobieństwa. Nośnikiem rozkładu prawdopodobieństwa nazywamy taki, najmniejszy i domknięty zbiór liczb $Supp \subset \mathbb{R}$, dla których $P(x \in Supp) = 1$ (ang. $\boldsymbol{supp}ort$). Ważną właściwością nośnika jest jego rozpiętość mierzona jako różnica skrajnych wartości należących do nośnika: $R = \max Supp - \min Supp$; (ang. $\boldsymbol{r}ange$). Wielkość tę można określić nie tylko wtedy, gdy rozkład jest ciągły, ale także wtedy gdy jest dyskretny lub mieszany.

Nasza wiedza aprioryczna na temat przedziału wartości należnego menzurandowi pochodzi bezpośrednio, lub – częściej – pośrednio z pomiarów wykonanych wcześniej. Pomiary związane z badaniami podstawowymi, wykonywane w warunkach niewielkiej wiedzy nt. obiektu i czynników wpływających na realizację badaną, stanowią niewielką część wszystkich wykonywanych pomiarów. Większość stanowią pomiary obiektów naturalnych, o których stosunkowo wiele już wiadomo, oraz pomiary obiektów będących wytworem inżynierskiej działalności człowieka, a więc - o dobrze znanej budowie i zasadzie działania. W obiektach tego typu granice możliwych wartości realizacji badanej są następstwem znanych właściwości użytych elementów, sposobu zasilania systemu, jego pobudzania, składu chemicznego, znanych wielkości wpływających itd.

Przykład 1: Z powodu licznych, fizycznych i biochemicznych, sprzężeń zwrotnych występujących w organizmach żywych (zwłaszcza tych wyżej zorganizowanych) parametry tych organizmów (np. skład i parametry płynów ustrojowych) zwykle mieszczą się w stosunkowo wąskich granicach. Odchylanie się wartości parametru od tych granic jest zwykle związane z pojawieniem się jakiejś patologii. Jeśli brak występowania określonej patologii może być stwierdzony na podstawie objawów towarzyszących, to zawieranie się w znanych granicach wartości wielu parametrów płynu jest niemal pewne.

Przykład 2: Właściwości złożonych układów elektronicznych, także analogowych, zależą od parametrów tworzących je elementów i układów scalonych. Użyteczność wyspecjalizowanych układów scalonych w danym zastosowaniu zależy zwykle od kilku kluczowych parametrów. Wartości tych parametrów producent podaje w karcie katalogowej układu i specjalnie o nie dba w toku produkcji wykonując stosowne pomiary międzyoperacyjne i końcowe. Zwykle istotne jest, by kontrolowany parametr w lub jego moduł spełniał jedną z następujących relacji: $w_{min} < w < w_{max}$ (np. napięcie referencyjne scalonego wzorca napięcia, rezystancja rezystora precyzyjnego), $w_{min} < w$ (np. wzmocnienie z otwartą pętlą wzmacniacza operacyjnego), lub $w < w_{max}$ (np. moduł błędu nieliniowości scalonego przetwornika a/c, moduł wejściowego napięcia niezrównoważenia wzmacniacza operacyjnego, itd.). Jeżeli w wyniku pomiarów kontrolnych producent stwierdza, że któreś z wymaganych ograniczeń nie jest spełnione, to takie elementy wadliwe typowo eliminuje się z serii. Niekiedy możliwe jest jeszcze doregulowanie parametru niespełniającego deklarowanych wymagań, często -niestety -kosztem wartości innych parametrów. Ostatecznie więc użytkownik otrzymuje element, w którym parametr ma rozkład o nośniku ograniczonym przynajmniej jednostronnie. Analiza histogramów dostępnych w niektórych kartach katalogowych pokazuje, że zarówno eliminacja elementów wadliwych jak i ew. doregulowywanie wartości parametru powodują, że wynikowy rozkład parametru jest nie tylko wyraźnie ucięty, ale często także nienaturalny, tzn. niepodobny do funkcji gęstości prawdopodobieństwa znanych z teorii. Ucięte rozkłady wartości parametrów elementów składowych powodują, że właściwości całego systemu, w którym takie elementy wykorzystano, też mają rozkłady ucięte (przynajmniej jednostronnie), co można stwierdzić w drodze analiz lub symulacji metodą Monte Carlo (MC).

Przykład 3: Poprawne wykonanie większości pomiarów precyzyjnych wymaga przybliżonej choćby znajomości wartości mierzonej [1]. Ten wstępny przedział dla wartości realizacji mierzonej uzyskuje się zwykle innym narzędziem pomiarowym, lub na podstawie wiedzy o obiekcie badanym. Dotyczy to nie tylko metody różnicowej, której racjonalne zaprojektowanie wymaga wstępnej znajomości wartości menzurandu, ale także metody zerowej. Np. kompensacyjny pomiar napięcia, zwłaszcza w przypadku stosowanych dawniej ogniw normalnych, wymagał wstępnej znajomości siły elektromotorycznej ogniwa, tak aby komparacja nie doprowadziła do nadmiernego jego obciążenia.

Przykład 4: Błąd segmentacji to błąd związany z pomiarami parametrów uśrednionych sygnałów (prawie-) okresowych spowodowany tym, że rzeczywisty czas uśredniania nie jest całkowitą krotnością okresu sygnału badanego. Właściwości probabilistyczne tego błędu są znane i mają one rozkład ograniczony, [2]. Nośnik tego rozkładu jest zależny od zakresu zmienności sygnału.

Jak więc widać ograniczoność rozkładu prawdopodobieństwa wielkości mierzonej jest zwykle następstwem wcześniej wykonanych pomiarów. Rzadko dziś z tej wiedzy korzystamy, zapewne z powodu trudności formalnie ścisłego jej uwzględniania w wynikach końcowych. Celem tej pracy jest pokazanie, w jaki sposób rozpiętość nośnika propaguje się w różnych rodzajach odwzorowań typowych w metrologii.

2. PROPAGACJA ROZPIĘTOŚCI NOŚNIKA W ODWZOROWANIACH FUNKCYJNYCH

W pomiarach pośrednich, a także w analizach oddziaływania na wynik pomiaru wielkości wpływających, trzeba rozważać zależności funkcyjne $y = h(x_1, x_2, .., x_m)$, w których y oraz $x_1, x_2, .., x_m$ oznaczają realizacje zmiennych losowych. Oczywiście, jeśli wszystkie rozkłady zmiennych losowych będących zmiennymi niezależnymi, mają ograniczone nośniki, a ponadto funkcja h nie ma osobliwości w wielowymiarowym prostopadłościanie opisanym przez te nośniki, to rozkład zmiennej losowej wyniku Y też ma rozkład o nośniku ograniczonym. Niekiedy nośnik zmiennej losowej

Y może być ograniczony także wtedy, gdy niektóre nośniki zmiennych losowych $X_i, i = 1,..,m$ nie są ograniczone, ale sama funkcja h przyjmuje wartości ograniczone na tych nieograniczonych zbiorach będących nośnikami. Postać szczegółową rozkładu zmiennej losowej Y wylicza się na ogół dość zawile i często prościej jest zrobić badania symulacyjne metodą MC, [3].

Postać funkcji opisującej rozkład zmiennej losowej Y stosunkowo prosto analizuje się w przypadku, gdy funkcja h ma postać sumy $Y = \sum_{i=1}^{m} X_i$, a zmienne losowe X_i są niezależne. Wtedy funkcja gęstości prawdopodobieństwa zmiennej Y jest splotem funkcji gęstości zmiennych składowych. Nietrudno wówczas wykazać, że gdy składniki mają nośniki ograniczone, to rozpiętość nośnika sumy wynosi: $R_Y = \sum_{i=1}^{m} R_{X_i}$, gdzie R_{X_i} oznaczają rozpiętości nośników poszczególnych zmiennych losowych X_i. Wynik ten oznacza, że choć rozpiętość nośnika skończonej sumy zmiennych losowych rośnie wraz z liczbą składników, to jednak nośnik sumy nadal jest ograniczony.

Interesujące jest zbadanie wartość ilorazu R_Y/σ_Y, w którym σ_Y oznacza odchylenie standardowe sumy. Dzięki niezależności zmiennych losowych można napisać $R_Y/\sigma_Y = \left(\sum_{i=1}^{m} R_{X_i}\right)/\sqrt{\sum_{i=1}^{m} {\sigma_{X_i}}^2}$. Wartość tego ilorazu można oszacować z dołu zakładając, że występujące w mianowniku wariancje zmiennych losowych X_i, mają największe możliwe wartości, przy danych rozpiętościach nośników. Takie graniczne wartości wariancji osiągają zmienne losowe o rozkładzie dwupunktowym, przyjmujące z równymi prawdopodobieństwami wyłącznie skrajne wartości swoich nośników. Wariancje takich zmiennych losowych mają wartości: ${\sigma_{X_i}}^2 = {R_{X_i}}^2/4$. Tak więc, badany iloraz musi mieć wartość $R_Y/\sigma_Y > 2\left(\sum_{i=1}^{m} R_{X_i}\right)/\sqrt{\sum_{i=1}^{m} {R_{X_i}}^2} > 2$. W ostatnim oszacowaniu skorzystano z nierówności trójkąta. Gdy liczba sumowanych zmiennych losowych rośnie nieograniczenie to: $\lim_{m\to\infty} R_Y/\sigma_Y = \infty$ co oznacza, że rozpiętość nośnika sumy, choć dla skończonych m zawsze jest ograniczona, to jednak rośnie szybciej niż odchylenie standardowe sumy. Ten ostatni wynik nie jest zaskakujący, gdyż jest on następstwem centralnego twierdzenia granicznego. Rozkład sumy zmiennych losowych dąży do rozkładu Gaussa, który ma nośnik nieograniczony, ale ograniczoną sigmę, więc ich iloraz dąży do nieskończoności.

Podobnie można zbadać rozpiętość nośnika wartości średniej m niezależnych zmiennych losowych o ograniczonych nośnikach. Warto w tym celu zauważyć, że jeśli $Y = a \cdot X + b$, gdzie $a, b \in \mathbb{R}$, to $R_Y = a \cdot R_X$. Dzięki temu staje się oczywiste, że dla średniej $Y = (1/m)\sum_{i=1}^{m} X_i$ mamy $R_Y = (1/m)\sum_{i=1}^{m} R_{X_i}$. Podobnie jak dla sumy: $\lim_{m\to\infty} R_Y/\sigma_Y = \infty$, sygnalizując że graniczny rozkład średniej jest normalny i ma nieograniczony nośnik.

3. MODEL BAYESOWSKI OGRANICZANIA NOŚNIKA

Przyjmijmy, że celem analiz jest rozkład prawdopodobieństwa zmiennej losowej *X*. Może to być np. rozkład opisujący prawdopodobne wartości wybranego parametru sygnału. Naszą wiedzę o wartości tej zmiennej losowej opisuje funkcja gęstości prawdopodobieństwa (ang. ***p**robability **d**ensity **f**unction*, pdf). Dostępna wiedza aprioryczna *WA* pozwala określić prawdopodobieństwa z jakim zmienna losowa *X* może przyjmować wartości z poszczególnych przedziałów osi liczbowej. Opisuje je warunkowa pdf $f_a(x|WA)$. Znak $|$ oznacza „zakładając, że …”. Zdobycie informacji dodatkowych *I* może umożliwić modyfikację rozkładu *a priori* $f_a(x|WA)$ do rozkładu *a posteriori* $f_p(x|WA \wedge I)$. Potrzebna jest do tego funkcja gęstości warunkowej $f_l(I|WA \wedge x)$ (funkcja wiarygodności, ang. ***l**ikelihood*), która wyraża związek wartości zmiennej losowej *X* z uzyskanymi informacjami na temat rozkładu. Twierdzenie Bayes’a można zapisać w postaci:

$$f_p(x|WA \wedge I) \propto f_a(x|WA) \cdot f_l(I|WA \wedge x), \tag{1}$$

gdzie znak $\propto$ oznacza proporcjonalność. Aby móc zastąpić proporcjonalność znakiem równości, całka z prawej strony (1) powinna być równa jedności. Taka normalizacja może być uzyskana przez podzielenie prawej strony (1) przez współczynnik równy: $C = \int_{\mathbb{R}} f_a(x|WA) \cdot f_l(I|WA \wedge x)\ dx$.

Jeżeli informacja *I* zdobyta o badanym rozkładzie orzeka jedynie to, że: $\underline{X} < X < \overline{X}$, tzn. że rozkład zmiennej losowej *X* ma ograniczony nośnik, to funkcja gęstości tej zmiennej losowej ma co najmniej dwa parametry: X_{min}.oraz X_{max} o wartościach, które znamy i równych odpowiednio $\underline{X}$ i $\overline{X}$. Wprawdzie informacja *I* nie orzeka niczego o zachowaniu się $f_l(I|WA \wedge x)$ pomiędzy podanymi wartościami, ale wiemy na pewno, że dla $x \notin (\underline{X};\ \overline{X})$ będzie: $f_l(I|WA \wedge x) \equiv 0$. To zaś oznacza, że całka w powyższym wzorze do obliczania współczynnika *C* jest zbieżna, a funkcja gęstości *a posteriori* $f_p(x|WA \wedge \underline{X} < X < \overline{X}) \equiv 0$ na zewnątrz odcinka $(\underline{X};\ \overline{X})$.

W literaturze dotyczącej analiz bayesowskich zwykle sugeruje się, by zerowym prawdopodobieństwem, lub jego zerową gęstością, opisywać wyłącznie takie zdarzenia, których pojawienie się jest logicznie wykluczone przy wiedzy, którą dysponujemy [4]. Z tego powodu funkcje gęstości wykorzystywane w tych analizach wprawdzie szybko opadają poza wartościami, które zmienna losowa przyjmuje najbardziej prawdopodobnie, ale też nie wykluczają zupełnie możliwości ich przyjęcia. Jeżeli wtedy zbiory, na których obie funkcje po prawej stronie (1) przyjmują wartości znacząco niezerowe, są rozłączne (co oznacza, że informacje dostępne *a priori* i te pochodzące z pomiarów i analiz są prawdopodobnie sprzeczne), to dla wszystkich wartości *x* ich iloczyn przyjmuje wartości niezerowe, choć bardzo małe. Jeżeli normowanie takiego iloczynu odbywa się bez zwracania uwagi na wartość współczynnika *C*, która jest bardzo duża w takich przypadkach, to nie mamy innych objawów prawdopodobnej sprzeczności. Posługiwanie się jawnie uciętymi funkcjami gęstości nie ma tej wady, bo rozłączność nośników funkcji gęstości *a priori* f_a i wiarygodności f_l prowadzi do zerowych wartości pdf *a posteriori* na całej osi liczbowej, sugerując powtórne zweryfikowanie wszystkich dostępnych informacji.

4. WPŁYW OGRANICZANIA NOŚNIKA NA KURTOZĘ I PRZEDZIAŁ POKRYCIA

Kurtoza: Jak wiadomo, kurtoza czyli unormowany czwarty moment centralny: μ_4/σ^4, jest miarą „tłustości" ogonów funkcji gęstości. Jeżeli dla niezmiennej postaci funkcji gęstości, kolejne zdobyte informacje $I_i,\ i = 1,2,..$ będą upoważniać do ucinania zewnętrznych części nośnika, tzn. nowy nośnik $Supp_{i+1}$ nie tylko jest podzbiorem starego $Supp_i$, ale także: $\max Supp_{i+1} < \max Supp_i$ oraz $\min Supp_{i+1} > \min Supp_i$ (czyli $R_{i+1} < R_i$), a także dodatkowo znormalizuje się uzyskany rozkład, to taki zabieg zawsze spowoduje zmniejszenie wartości zarówno μ_4 jak i wariancji σ^2. Ponieważ jednak μ_4 silniej niż σ^2 zależy od mas prawdopodobieństwa zlokalizowanych daleko od średniej, więc kurtoza zawsze będzie maleć w miarę opisanego zewnętrznego ograniczania nośnika rozkładu. Gdy ponadto granice rozkładu uciętego będą się do siebie zbliżać: $\min Supp \to \max Supp$, powodując że pdf w pozostawionej części nośnika wykazuje coraz mniejszą zmienność, to $\mu_4/\sigma^4 \to \frac{9}{5}$, czyli kurtoza dąży do wartości charakterystycznej dla rozkładu równomiernego.

Przedział pokrycia: Rozważamy wpływ ciągu kolejnych ograniczeń zewnętrznych nośnika jednomodalnej pdf na położenie symetrycznego (wg prawdopodobieństwa) przedziału niepewności $(x_l;\ x_u)$, dla zadanego prawdopodobieństwa pokrycia p_{cov} (ang. ***cov****erage probability*). Nietrudno sprawdzić, że dla ustalonego p_{cov}, im bardziej uda się ograniczyć od zewnątrz nośnik, co wymaga unormowania pozostałej części pdf, tym krańce przedziału niepewności stają się bliższe (od wewnątrz) znanym granicom nośnika. W przypadku, gdy przebieg ucinanej funkcji gęstości nie jest znany, wówczas znane granice nośnika mogą stanowić dobre oszacowania zewnętrzne granic przedziału niepewności: $\min Supp \lesssim x_l$ oraz $x_u \lesssim \max Supp$. W przypadku, gdy funkcja gęstości zmienia się łagodnie w pobliżu mody, zaś ucięcie nośnika jest już tak znaczące, że pozostałą jeszcze część pdf można uznać za prawie stałą na odcinku $(\min Supp\,;\ \max Supp)$, wówczas szerokość przedziału niepewności wynosi: $x_u - x_l \cong R \cdot p_{cov}$.

5. PODSUMOWANIE

Skuteczność analiz bayesowskich dotyczy zwłaszcza ich wersji parametrycznej, kiedy postać funkcyjna funkcji gęstości f_l jest znana (przynajmniej co do klasy rozkładów) a celem jest oszacowanie wartości parametrów rozkładu na podstawie próbek zmiennej losowej. W pomiarach postać funkcyjna pdf nie zawsze jest znana, a nadto nie zawsze mamy dostęp do wystarczająco wielu niezależnych próbek, aby można było korzystać z centralnego twierdzenia granicznego. Dlatego wiedzę dotyczącą granic nośnika należy skrupulatnie wykorzystywać zwłaszcza wtedy, gdy rozpiętość nośnika jest mniejsza od kilku odchyleń standardowych.

LITERATURA

1. Marcyniuk A.: Teoria pomiaru. Podstawy metrologicznej interpretacji wyniku pomiaru dla metrologów elektryków. Wyd. II, skrypt nr 1168 Politechniki Śląskiej, Gliwice, 1984.
2. Smołalski G.: Segmentation error in averaged parameters' measurements of periodic signals: its upper limits and general probabilistic properties. *Measurement*, 29 (2001), ss. 21-30.
3. Jakubiec J.: Błędy i niepewności danych w systemie pomiarowo-sterującym. Wydawnictwo Politechniki Śląskiej, Gliwice, 2010.
4. O'Hagan A.: Probability. Methods and measurement. Chapman and Hall, Londyn, 1988, rozdz. 3.4.