<u>Charakterystyka rozkładu *a posteriori*</u>

Przed nami kolejne zagadnienie – kluczowe w bayesowskiej analizie statystycznej: otóż, w jaki sposób *przestawić/komunikować* naszą końcową, ujętą w rozkładzie *a posteriori* wiedzę o parametrach? Odpowiedź: poprzez jego charakterystyki, *porównując* je z odpowiednimi charakterystykami rozkładu *a priori*. To porównanie rozkładu *a posteriori* z rozkładem *a priori* pozwala stwierdzić, na ile dane zmodyfikowały nasze wstępne przekonania/wyobrażenia o parametrach. Charakterystykami rozkładów *a priori* i *a posteriori*, które się z reguły poddaje analizie, są:

1) Gęstości rozkładów brzegowych
2) Charakterystyki położenia:
    a. Wartość oczekiwana
    b. Modalna
    c. Kwantyle (w tym mediana)
3) Charakterystyki rozproszenia:
    a. Wariancja
    b. Odchylenie standardowe
    c. Rozstęp międzykwartylowy
4) Współczynniki korelacji między parametrami
5) Prawdopodobieństwo interesującego podzbioru przestrzeni parametrów

W dalszym ciągu omówimy kolejno wymienione wyżej wielkości/obiekty.

<u>Ad 1) Gęstości rozkładów brzegowych</u>

Jedną z podstawowych zalet wnioskowania bayesowskiego jest to, że dostarcza ono pełnego obrazu *niepewności* związanej z estymacją parametrów, przy czym niepewność ta jest odzwierciedlona w stosownych *rozkładach prawdopodobieństwa* (*a priori* – przed uwzględnieniem informacji zawartych w danych, *a posteriori* – po uwzględnieniu tych informacji). O kształtowaniu się tej niepewności najpełniej informuje badacza przebieg wykresów odpowiednich funkcji gęstości (które mogą się charakteryzować mniej lub bardziej regularnym przebiegiem, odznaczać się wielomodalnością, grubymi ogonami, asymetriami, „garbami”, itp.). Trzeba jednak być świadomym naturalnego ograniczenia zdolności percepcyjnych naszych zmysłów już w przypadku rozkładów 3- i więcej-wymiarowych. Oczywiście, graficzna prezentacja funkcji gęstości parametrów jest możliwa tylko w przypadku, gdy wektor θ jest 1-, 2- lub maksymalnie 3-wymiarowy. Wówczas wykresy gęstości są, odpowiedni, tworzone w 2D, 3D i 4D. Ten ostatni przypadek – choć teoretycznie możliwy – wymaga już nie lada „skillu” w zakresie wyobraźni przestrzennej i odpowiednich narzędzi graficznych (o tworzeniu wykresów 4D mogą Państwo zobaczyć kilka prezentacji na YouTube;). Dlatego też, w praktyce, najczęściej prezentowane są wykresy 1-wymiarowych rozkładów brzegowych (wykresy 2D), rzadziej – 2-wymiarowych (wykresy 3D). W każdym z wymienionych tu przypadków koniecznym jest uzyskanie funkcji gęstości stosownego rozkładu *brzegowego*:

- *a priori*:
    - $p(\theta_i)$, gdy interesuje nas pojedynczy parametr, θ_i;
    - $p(\theta_i, \theta_j)$, gdy interesuje nas para (θ_i, θ_j);

- *a posteriori*:
  - $p(\theta_i|y)$, gdy interesuje nas pojedynczy parametr, θ_i;
  - $p(\theta_i, \theta_j|y)$, gdy interesuje nas para (θ_i, θ_j).

Zauważmy, że rozkłady: $p(\theta_i, \theta_j)$ i $p(\theta_i, \theta_j|y)$ są, z jednej strony, *łączne* (bo dotyczą dwóch parametrów jednocześnie), z drugiej zaś – *brzegowe*, ponieważ otrzymujemy je z rozkładów, odpowiednio: $p(\theta)$ i $p(\theta|y)$ poprzez scałkowanie tych ostatnich względem wszystkich pozostałych parametrów. Niemniej jednak, zwykle o $p(\theta_i, \theta_j)$ i $p(\theta_i, \theta_j|y)$ mówi się jako o rozkładach *brzegowych* (z dopowiedzeniem: 2-wymiarowych), ponieważ ten ich aspekt wydaje się bardziej wart podkreślenia.

W praktyce, zwykle dysponujemy *analitycznymi* (jawnymi) formułami (ang. *closed-form formulae/expressions*) gęstości tylko brzegowych rozkładów *a priori*, $p(\theta_i)$ i $p(\theta_i, \theta_j)$. Wobec tego graficzna ich prezentacja raczej nie nastręcza trudności (istnieje wiele pakietów graficznych, gdzie na podstawie analitycznego wzoru pewnej 1- lub 2-argumentowej funkcji można przedstawić jej wykres, odpowiednio, w 2D lub 3D). Nieco inaczej ma się sprawa z gęstościami rozkładów *a posteriori*, $p(\theta_i|y)$ i $p(\theta_i, \theta_j|y)$, gdyż rzadko kiedy gęstości tych rozkładów dają się wyrazić jawnymi wzorami. Na naszych zajęciach, owszem, zetkniemy się z takimi modelami, ale w dużej mierze będą one służyć raczej celom dydaktycznym; później zajmiemy się takimi modelami – dalece bardziej praktycznymi – w których wyznaczenie analitycznych formuł gęstości brzegowych rozkładów *a posteriori* jest już niemożliwe. To, z kolei, będzie stanowić naszą motywację do wykorzystania stosownych metod numerycznych (np. MCMC) w celu szacowania/aproksymacji charakterystyk tychże rozkładów. Zwykle produktem końcowym tych metod jest próba pseudolosowa z rozkładu *a posteriori* (w domyśle – łącznego, dla wszystkich parametrów, czyli $p(\theta|y)$), na podstawie której dają się wyznaczyć *histogramy*, które – uwaga – są *estymatorami* funkcji gęstości.[1] Za pomocą histogramów możemy aproksymować (niedające się wyznaczyć analitycznie) gęstości zarówno 1-, jak i 2-wymiarowych brzegowych rozkładów *a posteriori*. Do zobrazowania gęstości 1-wymiarowych brzegowych rozkładów *a posteriori*, faktycznie, z reguły wykorzystuje się histogramy (najlepiej z naniesionymi nań odpowiadających im gęstościami brzegowych rozkładów *a priori*, by umożliwić wzrokową ocenę tego, na ile dane modyfikują wstępne przekonania o parametrach). Choć w celu zobrazowania gęstości rozkładów 2-wymiarowych również można posłużyć się histogramem (zbudowanym w 3D), to jednak warto wziąć pod rozwagę także inne możliwości (tj. typy wykresów) ich wizualizacji, jak np. różne warianty wykresu poziomicowego (ang. *level plot*; czasem zwanego także wykresem konturowym, ang. *contour plot, contour map*), czy mapy ciepła (ang. *heat maps*; wartości funkcji gęstości reprezentowane są przez odpowiednie kolory). Na tych wykresach (2D!) – fundujących nam widok z lotu ptaka – można znacznie łatwiej dostrzec wszelakie modalne (w rozkładzie 2-wymiarowym) oraz związki korelacyjne i zależności pomiędzy danymi dwoma parametrami.

Ad 2) Charakterystyki położenia

Ad 2.a) Wartość oczekiwana – *a priori:* $E(\theta)$; *a posteriori*: $E(\theta|y)$

Pamiętajmy, że wartość oczekiwana wektora losowego to wektor wartości oczekiwanych poszczególnych jego składowych:

- *a priori*:

$$E(\theta) = [E(\theta_1)\ \ E(\theta_2)\ \ ...\ \ E(\theta_s)]', \tag{3}$$

[1] Histogramy nie są jedynymi dostępnymi (nieparametrycznymi) estymatorami funkcji gęstości. Inną klasę (także nieparametrycznych) estymatorów gęstości rozkładów prawdopodobieństwa stanowią tzw. *jądrowe estymatory gęstości* (ang. *kernel density estimators*), których wynikiem jest w pewnym stopniu wygładzony histogram. Estymatory jądrowe – choć nieco trudniejsze w implementacji – są już obecnie standardowym modułem licznych pakietów statystycznych.

gdzie s oznacza wymiar wektora θ;

- *a posteriori*:

$$E(\theta \mid y) = [E(\theta_1 \mid y) \;\; E(\theta_2 \mid y) \;\; \ldots \;\; E(\theta_s \mid y)]'. \tag{4}$$

Powyższa uwaga oznacza, że jeśli dysponujemy wartością oczekiwaną danego *łącznego* rozkładu wektora $\theta = [\theta_1, \theta_2, \ldots, \theta_s]'$, tj. $E(\theta)$ i $E(\theta|y)$, to poszczególne współrzędne $E(\theta)$ i $E(\theta|y)$ stanowią wartości oczekiwane *brzegowych* rozkładów poszczególnych współrzędnych wektora θ. Przykładowo, druga współrzędna w zapisie (3) jest wartością oczekiwaną *a priori* parametru θ_2 zarówno w rozkładzie *łącznym* (dla całego θ), jak i *brzegowym* (dla tegoż θ_2).

Dodajmy jeszcze, że wartość oczekiwaną danego rozkładu brzegowego jest sens raportować tylko w sytuacji, gdy moment ten *istnieje*, tzn.

- *a priori*:

$$\int_{\Theta_i} \theta_i p(\theta_i) d\theta_i \neq \pm\infty$$ (nieco bardziej formalnie: $\int_{\Theta_i} |\theta_i| \, p(\theta_i) d\theta_i < +\infty$),

- *a posteriori*:

$$\int_{\Theta_i} \theta_i p(\theta_i \mid y) d\theta_i \neq \pm\infty$$ (nieco bardziej formalnie: $\int_{\Theta_i} |\theta_i| \, p(\theta_i \mid y) d\theta_i < +\infty$).

Jeśli rozkład *a posteriori* jest którymś ze znanych rozkładów prawdopodobieństwa, wówczas wiadomo, czy istnieje jego wartość oczekiwana, czy też nie (jak ma to miejsce np. w rozkładzie Cauchy'ego, czyli rozkładzie *t*-Studenta o 1 stopniu swobody). Niestety, w praktyce częstokroć rozkłady *a posteriori* nie należą do żadnych z poznanych dotąd rodzin rozkładów prawdopodobieństwa. W takich wypadkach zachodzi potrzeba *numerycznego* wyznaczania (aproksymacji) wszelakich charakterystyk rozkładu *a posteriori* poprzez wcześniejsze uzyskanie próby pseudolosowej z tego rozkładu (np. za pomocą metod MC-IS czy MCMC), a następnie obliczenie stosownych momentów *próbkowych*, np. średniej arytmetycznej z wygenerowanej próby jako oszacowania prawdziwej (niedającej się wyznaczyć analitycznie) wartości oczekiwanej rozkładu $p(\theta_i|y)$. Powstaje jednak wtedy pytanie: czy stosowny moment brzegowego rozkładu *a posteriori* w ogóle istnieje? Jest to o tyle ważne, że średnią arytmetyczną z próby możemy policzyć zawsze. Jednak w sytuacji, gdy prawdziwa wartość oczekiwana rozkładu *a posteriori* nie daje się wyznaczyć analitycznie (tudzież nie można analitycznie dowieść jej istnienia), wówczas nie możemy mieć pewności, że ów średnia arytmetyczna faktycznie pełni funkcję *oceny* wartości oczekiwanej (gdyż, powtórzmy, nie wiemy, czy ta ostatnia w ogóle istnieje). Zauważmy jednak, że bez względu na to, czy dany prawdziwy moment rozkładu istnieje czy nie, *próbkowe* momenty możemy zawsze wykorzystywać w charakterze zwykłych „opisowych" mierników, charakteryzujących próbę wygenerowaną z danego rozkładu; wówczas nie musimy się przejmować, czy prawdziwy moment istnieje, czy nie, ponieważ celem *nie* jest *oszacowanie* tegoż momentu poprzez odpowiedni moment próbkowy, a jedynie charakteryzacja (opis) próby pseudolosowej wygenerowanej z danego rozkładu. Oczywiście, powyższa uwaga dotyczy momentów dowolnego rzędu, a nie tylko wartości oczekiwanych.

Ad 2.b) Modalne 1-wymiarowych rozkładów brzegowych[2] – *a priori*: $Mo(\theta_i)$; *a posteriori*: $Mo(\theta_i|y)$

Przypomnijmy, że modalna pewnej zmiennej losowej X (skalarnej lub wektorowej) to taka jej wartość (lub wektor wartości – gdy X jest wektorem losowym), dla której funkcja gęstości (gdy X jest ciągłą zmienną losową) lub funkcja masy prawdopodobieństwa (gdy X jest dyskretna) osiąga maksimum (lokalne).[3] Tak więc:

– w rozkładach 1-wymiarowych (θ_i jest skalarem, więc modalna także jest wielkością skalarną):

$$Mo(\theta_i) = \arg\max_{\theta_i \in \Theta_i} p(\theta_i), \qquad Mo(\theta_i \mid y) = \arg\max_{\theta_i \in \Theta_i} p(\theta_i \mid y);$$

– w rozkładach wielowymiarowych (modalna rozkładu wektora $\theta = [\theta_1, \theta_2, \ldots, \theta_s]'$ będzie wielkością wektorową):

$$Mo(\theta) = \arg\max_{\theta \in \Theta} p(\theta), \qquad Mo(\theta \mid y) = \arg\max_{\theta \in \Theta} p(\theta \mid y).$$

Należy pamiętać, że, w ogólności, modalna danego rozkładu łącznego *nie* jest tożsama z wektorem modalnych jego rozkładów brzegowych, tj.

$$Mo(\theta) \neq [Mo(\theta_1)\ \ Mo(\theta_2)\ \ldots\ Mo(\theta_s)]', \qquad Mo(\theta \mid y) \neq [Mo(\theta_1 \mid y)\ \ Mo(\theta_2 \mid y)\ \ldots\ Mo(\theta_s \mid y)]',$$

choć są pewne wyjątki (np. wielowymiarowe rozkłady normalne i *t*-Studenta, czy, nieco ogólniej, wielowymiarowe rozkłady elipsoidalne).

W praktyce raczej rzadko wyznacza się (mniej lub bardziej dokładne) wartości modalnych w rozkładzie *a posteriori*. Z reguły konieczne jest w tym celu wykorzystanie zaawansowanych (zwłaszcza dla modalnych rozkładów wielowymiarowych) metod numerycznych, a pożytek z tego – raczej niewielki, tym bardziej, że o lokalizacji modalnej (czy modalnych – w przypadku rozkładów z wieloma modami) danego rozkładu najłatwiej (choć tylko w sposób przybliżony) można się dowiedzieć z wizualnej analizy jego funkcji gęstości (o czym poniżej; patrz podsekcja 2.g).

Ad 2.c) Kwantyle jednowymiarowych rozkładów brzegowych – *a priori*: $Q_\alpha(\theta_i)$; *a posteriori*: $Q_\alpha(\theta_i|y)$ (α oznacza rząd kwantyla)

Wyznaczanie kwantyli wielowymiarowego rozkładu prawdopodobieństwa, podobnie jak jego modalnej, zdecydowanie nie należy do zadań łatwych (choć, trzeba przyznać, że jest użyteczne na gruncie wielowymiarowej analizy danych). Dlatego też, w charakteryzacji rozkładu *a posteriori* z wykorzystaniem kwantyli zwykle ograniczamy się do raportowania kwantyli jego 1-wymiarowych rozkładów brzegowych. Kwantyl rzędu α rozkładu $p(\theta_i|y)$, czyli $Q_\alpha(\theta_i|y)$, to taka wartość parametru θ_i, że

$$\Pr\big(\theta_i < Q_\alpha(\theta_i \mid y) \big| y\big) = \int_{-\infty}^{Q_\alpha(\theta_i|y)} p(\theta_i \mid y) = \alpha .$$

Odpowiedni kwantyl rozkładu *a priori* definiujemy analogicznie:

---

[2] Oczywiście, modalną można rozważać – ogólnie – dla rozkładów wielo-, a niekoniecznie tylko 1-wymiarowych. Jednakże, w niniejszym opracowaniu główny akcent położono na przypadek tych ostatnich, gdyż w praktyce, analizę (zwykle wielowymiarowego) rozkładu *a posteriori* przeprowadza się częstokroć poprzez badanie charakterystyk właśnie 1-wymiarowych brzegowych rozkładów *a posteriori*, uzyskanych (poprzez stosowne całki) z rozkładu łącznego (rzadziej natomiast, 2-wymiarowych rozkładów brzegowych, choć i te łatwo dają się przedstawić na wykresach 3D lub za pomocą poziomic czy tzw. „heat maps").

[3] W rozkładach jednomodalnych ów maksimum jest jedyne i dlatego stanowi zarazem maksimum globalne. W rozkładach wielomodalnych kwestia jedyności (ang. *uniqueness*) maksimum jest nieco bardziej subtelna i nie będzie tu omawiana.

$$\Pr\left(\theta_i < Q_\alpha(\theta_i)\right) = \int_{-\infty}^{Q_\alpha(\theta_i)} p(\theta_i) = \alpha \, .$$

W praktyce, rozkład *a posteriori* zwykle nie jest żadnym ze znanych rozkładów prawdopodobieństwa, wobec czego kwantyle jego rozkładów brzegowych wyznacza się z próby pseudolosowej wygenerowanej z łącznego rozkładu *a posteriori*.

Oczywiście, wśród całej, nieprzeliczalnej (z punktu widzenia doboru rzędu kwantyla) rodziny kwantyli danego rozkładu, szczególną rolę odgrywa mediana. Jeśli dany (1-wymiarowy) rozkład prawdopodobieństwa jest rozkładem symetrycznym i istnieje jego wartość oczekiwana, to mediana i owa wartość oczekiwana są równe (jak w rozkładzie normalnym, czy w rozkładzie *t*-Studenta o liczbie stopni swobody większej od 1). W praktyce medianę danego brzegowego rozkładu *a posteriori* zwykle raportuje się wtedy, gdy rozkład ten jest asymetryczny (zwracając przy okazji uwagę, jak ma się mediana w porównaniu z wartością oczekiwaną).

Ad 3) Charakterystyki rozproszenia

Ad 3.a-b) Wariancje i odchylenia standardowe

→ Oznaczenia i wzory definicyjne:

- Wariancja pojedynczego parametru:
  - a priori: $Var(\theta_i) = E\left(\theta_i - E(\theta_i)\right)^2 = E\left(\theta_i^2\right) - E(\theta_i)^2$
  - a posteriori: $Var(\theta_i|y) = E\left[\left(\theta_i - E(\theta_i|y)\right)\middle|y\right]^2 = E\left(\theta_i^2|y\right) - E(\theta_i|y)^2$
- Odchylenie standardowe pojedynczego parametru:
  - a priori: $D(\theta_i) = \sqrt{Var(\theta_i)}$
  - a posteriori: $D(\theta_i|y) = \sqrt{Var(\theta_i|y)}$

Warto pamiętać, że – podobnie jak w przypadku wartości oczekiwanej – wariancja (a tym samym odchylenie standardowe) danego rozkładu może nie istnieć, co wynika z faktu, że jest to wielkość definiowana jako stosowna całka oznaczona (a ta może nie być zbieżna), po całej osi liczb rzeczywistych. Jeśli dany rozkład jest którymś ze znanych rozkładów prawdopodobieństwa, to znając jego własności możemy sprawdzić, czy istnieje drugi moment (i, ewentualnie, przy jakich warunkach). W przeciwnym razie, tj. gdy dany rozkład nie jest „standardowy” – jak ma to często miejsce w przypadku rozkładów *a posteriori* w bardziej zaawansowanych modelach ekonometrycznych – odchylenie standardowe rozkładu przybliżamy (szacujemy) za pomocą *próbkowego* odchylenia standardowego, obliczonego na podstawie próby pseudolosowej wygenerowanej z tegoż rozkładu. Wówczas, gdy chcemy traktować tak obliczoną wielkość jako *oszacowanie* prawdziwego, nie dającego się analitycznie wyznaczyć odchylenia standardowego, musimy być świadomi, iż tylko zakładamy, że ów nieznane odchylenie standardowe *istnieje* (owo „istnienie” z reguły także pozostaje w sferze *założeń*, gdyż nie da się go dowieść analitycznie dla rozkładów *a posteriori* o „niestandardowej” postaci).

Ad 3.c) Rozstęp międzykwartylowy

Do charakterystyki brzegowych rozkładów *a priori* i *a posteriori* pod względem ich *rozproszenia* najczęściej wykorzystuje się odchylenia standardowe (rzadziej wariancje), choć, naturalnie, nic nie stoi na przeszkodzie, aby w tym samym kontekście wykorzystać jeszcze inne mierniki rozproszenia, np. te oparte na kwartylach (znane ze statystyki opisowej), w tym:

- rozstęp międzykwartylowy (ang. *interquartile range*, IQR):
  - *a priori*: $IQR(\theta_i) = Q_{0,75}(\theta_i) - Q_{0,25}(\theta_i)$
  - *a posteriori*: $IQR(\theta_i \mid y) = Q_{0,75}(\theta_i \mid y) - Q_{0,25}(\theta_i \mid y)$
- odchylenie ćwiartkowe (ang. *midhinge*, *the 25% trimmed mid-range*) – równe połowie rozstępu międzykwartylowego.

Ad 4) Macierz korelacji wektora parametrów – *a priori*: *R*(θ); *a posteriori*: *R*(θ|*y*)

Zapiszmy „porządne" wzory:

- Macierz korelacji *a priori* wektora parametrów:

$$R(\theta) = B(\theta)V(\theta)B(\theta) = \left[\frac{Cov(\theta_i,\theta_j)}{\sqrt{Var(\theta_i)Var(\theta_j)}}\right]_{i,j=1,2,\dots,s}$$

gdzie $V(\theta)$ jest macierzą kowariancji *a priori* wektora parametrów:

$$V(\theta) = E\left[\big(\theta - E(\theta)\big)\big(\theta - E(\theta)\big)'\right] = E(\theta\theta') - E(\theta)E(\theta)'$$

$$= \begin{bmatrix} Var(\theta_1) & Cov(\theta_1,\theta_2) & \dots & Cov(\theta_1,\theta_s) \\ Cov(\theta_2,\theta_1) & Var(\theta_2) & \dots & Cov(\theta_2,\theta_s) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\theta_s,\theta_1) & Cov(\theta_s,\theta_2) & \dots & Var(\theta_s) \end{bmatrix},$$

natomiast $B(\theta)$ jest macierzą diagonalną z odwrotnościami odchyleń standardowych *a priori* na przekątnej:

$$B(\theta) = \begin{bmatrix} \big(Var(\theta_1)\big)^{-0,5} & 0 & \dots & 0 \\ 0 & \big(Var(\theta_2)\big)^{-0,5} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \big(Var(\theta_s)\big)^{-0,5} \end{bmatrix}.$$

- Macierz korelacji *a posteriori* wektora parametrów:

$$R(\theta|y) = B(\theta|y)V(\theta|y)B(\theta|y) = \left[\frac{Cov(\theta_i,\theta_j|y)}{\sqrt{Var(\theta_i|y)Var(\theta_j|y)}}\right]_{i,j=1,2,\dots,s}$$

gdzie $V(\theta|y)$ jest macierzą kowariancji *a posteriori* wektora parametrów:

$$V(\theta|y) = E\left[\big(\theta - E(\theta|y)\big)\big(\theta - E(\theta|y)\big)'|y\right] = E(\theta\theta'|y) - E(\theta|y)E(\theta|y)'$$

$$= \begin{bmatrix} Var(\theta_1|y) & Cov(\theta_1,\theta_2|y) & \dots & Cov(\theta_1,\theta_s|y) \\ Cov(\theta_2,\theta_1|y) & Var(\theta_2|y) & \dots & Cov(\theta_2,\theta_s|y) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\theta_s,\theta_1|y) & Cov(\theta_s,\theta_2|y) & \dots & Var(\theta_s|y) \end{bmatrix},$$

natomiast $B(\theta|y)$ jest macierzą diagonalną z odwrotnościami odchyleń standardowych *a posteriori* na przekątnej:

$$B(\theta|y) = \begin{bmatrix} \left(Var(\theta_1|y)\right)^{-0,5} & 0 & \dots & 0 \\ 0 & \left(Var(\theta_2|y)\right)^{-0,5} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \left(Var(\theta_s|y)\right)^{-0,5} \end{bmatrix}.$$

Na koniec przypomnijmy jeszcze, że elementy macierzy korelacji są współczynnikami korelacji *liniowej* (i tylko liniowej!). Uwaga ta jest o tyle ważna, że – zwłaszcza w rozkładach *a posteriori* – pomiędzy daną parą parametrów może zachodzić *nieliniowy* związek korelacyjny. Dlatego też, jeśli interesuje nas korelacja *a posteriori* pomiędzy danymi dwoma parametrami, np. θ_i i θ_j, wówczas dobrym pomysłem wydaje się wizualna analiza funkcji gęstości ich rozkładu brzegowego, tj. $p(\theta_i, \theta_j|y)$ (o czym szerzej w podsekcji 2.g). Naturalnie, można także próbować mierzyć ową korelację z wykorzystaniem przeznaczonych do tego – choć raczej rzadko wykorzystywanych w praktyce – mierników (zob. np. [Wang Q., Shen Y., Zhang J.Q., 2005, *A nonlinear correlation measure for multivariable data set*, Physica D, vol. 200]).

Ad 5) Prawdopodobieństwo interesującego podzbioru przestrzeni parametrów

W niektórych sytuacjach przedmiotem badania może być wnioskowanie o tym, na ile możliwe jest, że wartości interesujących nas parametrów mieszczą się w pewnym zbiorze, oznaczanym w dalszym ciągu symbolem *A*. Może to dotyczyć zarówno całego wektora parametrów (wówczas *A* jest podzbiorem „całej" przestrzeni parametrów Θ, tj. $\theta \in A \subseteq \Theta$), jego pojedynczych współrzędnych ($\theta_i \in A_i \subseteq \Theta_i$), lub – w przypadkach pośrednich – dowolnego podwektora wektora θ. Dla zachowania zwięzłości, poniższe rozważania dotyczą tylko dwóch pierwszych wariantów (koncepcyjnie, ten ostatni, tj. wariant „pośredni", nie różni się znacząco od tych dwóch wcześniejszych). Przykładem analizy, w której możemy mieć do czynienia z zarysowanym tu zagadnieniem, może być ciąg niezależnych prób Bernoulliego jako model opisujący kolejne rzuty monetą. W tym zagadnieniu (skalarny) parametr $\theta \in \Theta = (0, 1)$ możemy utożsamiać z prawdopodobieństwem „orła". Po wykonaniu serii rzutów (których wyniki: zera i jedynki, stanowią próbę i zapisane są w postaci wektora *y*) może nas interesować odpowiedź na pytanie: jakie jest prawdopodobieństwo *a posteriori*, że prawdopodobieństwo wyrzucenia orła jest większe niż 0,5? Zatem $A = (0{,}5,\ 1) \subset \Theta$. Czujemy, że jeśli tylko moneta jest „sprawiedliwa", wówczas powinna zachodzić równość: $\Pr(\theta \in A|y) = \Pr(\theta \notin A|y) = \Pr(\theta \in (0,\ 0{,}5)|y) = 0{,}5$. Rzecz w tym, że trzeba to sprawdzić, obliczając właśnie $\Pr(\theta \in A|y)$.

Poniżej przedstawiono pewne ogólne zapisy związane z wyznaczaniem prawdopodobieństwa interesującego nas zbioru wartości parametrów:

- *a priori*:
  - w rozkładzie łącznym (dla całego wektora parametrów): $\Pr(\theta \in A)$, gdzie $A \subseteq \Theta$

$$\Pr(\theta \in A) = \int_A p(\theta) d\theta\,;$$

    (Zwróćmy uwagę, że powyższa całka jest wielokrotna, a dokładniej: *s*-krotna.)

  - w rozkładzie brzegowym (dla pojedynczego parametru): $\Pr(\theta_i \in A_i)$, gdzie $A_i \subseteq \Theta_i$

$$\Pr(\theta_i \in A_i) = \int_{A_i} p(\theta_i) d\theta_i\,;$$

    (Ta, z kolei, jest całką 1-krotną.)

- *a posteriori*:
  - w rozkładzie łącznym (dla całego wektora parametrów): $\Pr(\theta \in A|y)$, gdzie $A \subseteq \Theta$

$$\Pr(\theta \in A \mid y) = \int_A p(\theta \mid y) d\theta\,;$$

  - w rozkładzie brzegowym (dla pojedynczego parametru): $\Pr(\theta_i \in A_i|y)$, gdzie $A_i \subseteq \Theta_i$

$$\Pr(\theta_i \in A_i \mid y) = \int_{A_i} p(\theta_i \mid y) d\theta_i\,.$$

<u>Uwaga dot. prawdopodobieństwa w rozkładach łącznych</u>

Jeśli parametry danego rozkładu są niezależne i zbiór *A* jest iloczynem kartezjańskim zbiorów po wszystkich współrzędnych wektora θ, tzn. $A = \bigotimes_{j=1}^{s} A_j$, to prawdopodobieństwo zdarzenia $\theta \in A$ można obliczyć jako *iloczyn* odpowiednich całek jednowymiarowych z gęstościami *brzegowych* rozkładów parametrów θ_i jako funkcjami podcałkowymi, tj.

- *a priori*:

$$\Pr(\theta \in A) = \prod_{j=1}^{s} \Pr(\theta_j \in A_j) = \prod_{j=1}^{s} \int_{A_j} p(\theta_j) d\theta_j \ ;$$

- *a posteriori*:

$$\Pr(\theta \in A \mid y) = \prod_{j=1}^{s} \Pr(\theta_j \in A_j \mid y) = \prod_{j=1}^{s} \int_{A_j} p(\theta_j \mid y) d\theta_j \ .$$

<u>Uwaga dot. prawdopodobieństwa w rozkładach brzegowych</u>

Jeśli zbiór A_i jest przedziałem (na osi $\mathbb{R}$), tzn. $A_i = (a_i, b_i)$, to:

- *a priori*:

$$\Pr(\theta_i \in A_i) = \int_{a_i}^{b_i} p(\theta_i) d\theta_i = F_{\theta_i}(b_i) - F_{\theta_i}(a_i) \,,$$

gdzie $F_{\theta_i}(x)$ oznacza dystrybuantę brzegowego rozkładu *a priori* parametru θ_i;

- *a posteriori*:

$$\Pr(\theta_i \in A_i \mid y) = \int_{a_i}^{b_i} p(\theta_i \mid y) d\theta_i = F_{\theta_i \mid y}(b_i) - F_{\theta_i \mid y}(a_i) \,,$$

gdzie $F_{\theta_i \mid y}(x)$ oznacza dystrybuantę brzegowego rozkładu *a posteriori* parametru θ_i.