Optymalność wnioskowania bayesowskiego dla algorytmicznie losowych parametrów

Łukasz Dębowski
ldebowsk@ipipan.waw.pl

Instytut Podstaw Informatyki PAN
Warszawa

Główne idee

Algorytmiczna teoria informacji sugeruje atrakcyjną interpretację wnioskowania bayesowskiego:

1. Jeżeli wierzymy, że parametr jest typowy dla jakiegoś rozkładu a priori, może to oznaczać, że wierzymy, że parametr ów jest algorytmicznie losowy względem rozkładu a priori.
2. Można pokazać, że jeżeli parametr jest efektywnie estymowalny, to wnioskowanie bayesowskie jest optymalne wtedy i tylko wtedy, gdy parametr jest algorytmicznie losowy względem rozkładu a priori.

Bezprzedrostkowa maszyna Turinga (Chaitin 1975)

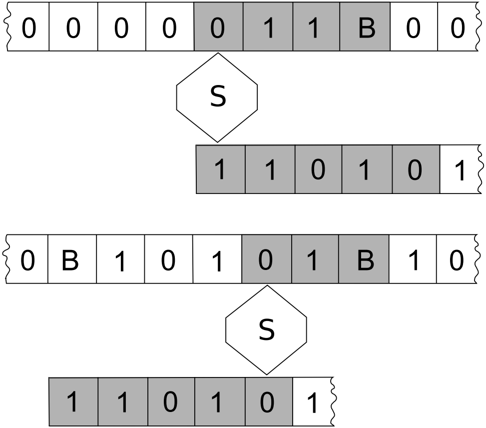
Rozpatrzmy maszynę, która ma dwie taśmy:

1. dwukierunkową taśmę $(\mathbf{X_i})_{\mathbf{i}\in\mathbb{Z}}$ wypełnioną symbolami **0**, **1** i **B**,
2. jednokierunkową taśmę $(\mathbf{Y_k})_{\mathbf{k}\in\mathbb{N}}$ wypełnioną symbolami **0** i **1**.

Głowica maszyny przesuwa się w obu kierunkach taśmy $(\mathbf{X_i})_{\mathbf{i}\in\mathbb{Z}}$ i tylko w kierunku rosnących **k** wzdłuż taśmy $(\mathbf{Y_k})_{\mathbf{k}\in\mathbb{N}}$.

Taśma $(\mathbf{X_i})_{\mathbf{i}\in\mathbb{Z}}$ może być czytana i nadpisywana,
taśma $(\mathbf{Y_k})_{\mathbf{k}\in\mathbb{N}}$ jest tylko do odczytu.

Bezprzedrostkowa maszyna Turinga



Początkowy i końcowy stan maszyny **S** dla $\mathbf{S(11010|011) = 01}$.

Zatrzymywanie się na wejściu

Oznaczmy $\mathbb{B} = \{0, 1\}^*$. Mówimy, że maszyna $\mathbf{S}$ zatrzymuje się na wejściu $(\mathbf{p}, \mathbf{q}) \in \mathbb{B}^* \times (\mathbb{B}^* \cup \mathbb{B}^{\mathbb{N}})$ i zwraca $\mathbf{w} \in \mathbb{B}^*$, jeżeli:

(A) Głowica w stanie początkowym odczytuje symbole $\mathbf{Y_1}$ i $\mathbf{X_1}$, a początkowy stan taśm to $\mathbf{Y_1^{|p|}} = \mathbf{p}$ oraz $\mathbf{X_1^{|q|+1}} = \mathbf{qB}$ lub $\mathbf{X_1^{\infty}} = \mathbf{q}$, jeżeli $\mathbf{q}$ jest ciągiem nieskończonym. Kładzie się także $\mathbf{X_i} = \mathbf{0}$ dla $\mathbf{i} < \mathbf{1}$ oraz $\mathbf{i} > |\mathbf{q}| + \mathbf{1}$, jeżeli $\mathbf{q}$ jest ciągiem skończonym (napisem).

(B) Głowica w stanie końcowym odczytuje symbole $\mathbf{Y_{|p|}}$ oraz $\mathbf{X_j}$, a końcowy stan taśmy $(\mathbf{X_i})_{\mathbf{i} \in \mathbb{Z}}$ to $\mathbf{X_j^{|w|+j}} = \mathbf{wB}$.

Zakładając, że (A) jest spełnione, piszemy

$$\mathbf{S(p|q)} = \begin{cases} \mathbf{w}, & \text{jeżeli (B) jest spełnione,} \\ \infty, & \text{jeżeli (B) nie jest spełnione dla żadnego } \mathbf{w} \in \mathbb{B}^*. \end{cases}$$

Dla danego $\mathbf{q}$ zbiór napisów $\mathbf{p}$ takich, że maszyna $\mathbf{S}$ zatrzymuje się na wejściu $(\mathbf{p}, \mathbf{q})$ jest zbiorem bezprzedrostkowym. Takie napisy nazywamy *samoograniczającymi się programami*.

Złożoność bezprzedrostkowa

Definicja (złożoność bezprzedrostkowa)

Warunkowa bezprzedrostkowa złożoność Kołmogorowa $\mathbf{K_S(w|q)}$ napisu $\mathbf{w} \in \mathbb{B}^*$ względem ciągu $\mathbf{q} \in (\mathbb{B}^* \cup \mathbb{B}^{\mathbb{N}})$ i maszyny $\mathbf{S}$ zdefiniowana jest jako

$$\mathbf{K_S(w|q)} := \min_{\mathbf{p} \in \mathbb{B}^*} \{|\mathbf{p}| : \mathbf{S(p|q) = w}\} . \tag{1}$$

Bezwarunkowa bezprzedrostkowa złożoność Kołmogorowa $\mathbf{K_S(w)}$ napisu $\mathbf{w} \in \mathbb{B}^*$ względem maszyny $\mathbf{S}$ zdefiniowana jest jako

$$\mathbf{K_S(w)} := \mathbf{K_S(w|\lambda)}, \tag{2}$$

gdzie $\boldsymbol{\lambda}$ jest napisem pustym.

Maszyny uniwersalne

Definicja (maszyna uniwersalna)

Maszynę $\mathbf{V}$ nazywa się *uniwersalną*, jeżeli dla każdej maszyny $\mathbf{S}$ istnieje napis $\mathbf{u} \in \mathbb{B}^*$ taki, że

$$\mathbf{V(up|q) = S(p|q)} \tag{3}$$

dla wszystkich $\mathbf{p} \in \mathbb{B}^*$ i $\mathbf{q} \in (\mathbb{B}^* \cup \mathbb{B}^{\mathbb{N}})$.

Uniwersalne maszyny istnieją. Przyjmujemy ten fakt bez dowodu.

Twierdzenie o niezmienniczości

Twierdzenie

Dla dowolnych dwu maszyn uniwersalnych $\mathbf{V}$ and $\mathbf{V}'$ istnieje stała $\mathbf{c}$ taka, że

$$|\mathbf{K}_{\mathbf{V}}(\mathbf{w}|\mathbf{q}) - \mathbf{K}_{\mathbf{V}'}(\mathbf{w}|\mathbf{q})| \leq \mathbf{c} \tag{4}$$

dla wszystkich $\mathbf{w} \in \mathbb{B}^*$ i $\mathbf{q} \in (\mathbb{B}^* \cup \mathbb{B}^{\mathbb{N}})$.

Dowód.

Mamy $\mathbf{V}(\mathbf{up}|\mathbf{q}) = \mathbf{V}'(\mathbf{p}|\mathbf{q})$ i $\mathbf{V}'(\mathbf{u}'\mathbf{p}|\mathbf{q}) = \mathbf{V}(\mathbf{p})$ dla pewnych napisów $\mathbf{u}$ i $\mathbf{u}'$. Stąd $\mathbf{K}_{\mathbf{V}}(\mathbf{w}|\mathbf{q}) \leq \mathbf{K}_{\mathbf{V}'}(\mathbf{w}|\mathbf{q}) + |\mathbf{u}|$ oraz $\mathbf{K}_{\mathbf{V}'}(\mathbf{w}|\mathbf{q}) \leq \mathbf{K}_{\mathbf{V}}(\mathbf{w}|\mathbf{q}) + |\mathbf{u}'|$. □

Złożoność bezprzedrostkowa II

Definicja (złożoność bezprzedrostkowa II)

Niech $\mathbf{V}$ będzie pewną maszyną uniwersalną. Kładziemy

$$\mathbf{K}(\mathbf{w}|\mathbf{q}) := \mathbf{K}_{\mathbf{V}}(\mathbf{w}|\mathbf{q}), \tag{5}$$

$$\mathbf{K}(\mathbf{w}) := \mathbf{K}_{\mathbf{V}}(\mathbf{w}). \tag{6}$$

Obliczalne funkcje dyskretne

Oznaczmy $\mathbb{A} = \mathbb{B}^* \cup \mathbb{B}^{\mathbb{N}}$. Złożoność Kołmogorowa dyskretnej funkcji częściowej $\mathbf{f} : \mathbb{A} \to \mathbb{B}^* \cup \{\infty\}$ definiujemy jako

$$\mathbf{K}(\mathbf{f}) := \min_{\mathbf{p}} \{|\mathbf{p}| : \forall_{\mathbf{u} \in \mathbb{A}} \mathbf{V}(\mathbf{p}|\mathbf{u}) = \mathbf{f}(\mathbf{u})\} . \tag{7}$$

Definicja (obliczalna funkcja dyskretna)

Funkcja $\mathbf{f} : \mathbb{A} \to \mathbb{B}^* \cup \{\infty\}$ jest *obliczalna*, jeżeli $\mathbf{K}(\mathbf{f}) < \infty$.

To uogólnia się na funkcje liczbowe, kodując w ustalony sposób liczby wymierne $\mathbb{Q}^* \leftrightarrow \mathbb{B}^*$ oraz rzeczywiste $\mathbb{R}^{\mathbb{N}} \leftrightarrow \mathbb{B}^{\mathbb{N}}$.

Semiobliczalne funkcje rzeczywiste

Definicja (semiobliczalna od dołu funkcja rzeczywista)

Funkcja $\mathbf{f} : \mathbb{A} \to \mathbb{R}$ nazywa się *semiobliczalną od dołu*, jeżeli istnieje obliczalna funkcja $\mathbf{A} : \mathbb{A} \times \mathbb{N} \to \mathbb{Q}$, która spełnia $\mathbf{A}(\mathbf{x}, \mathbf{k} + 1) \geq \mathbf{A}(\mathbf{x}, \mathbf{k})$ oraz

$$\forall_{\mathbf{x} \in \mathbb{A}} \lim_{\mathbf{k} \to \infty} \mathbf{A}(\mathbf{x}, \mathbf{k}) = \mathbf{f}(\mathbf{x}). \tag{8}$$

Definicja (semiobliczalna od góry funkcja rzeczywista)

Funkcja $\mathbf{f} : \mathbb{A} \to \mathbb{R}$ nazywa się *semiobliczalną od góry*, jeżeli istnieje obliczalna funkcja $\mathbf{A} : \mathbb{A} \times \mathbb{N} \to \mathbb{Q}$, która spełnia $\mathbf{A}(\mathbf{x}, \mathbf{k} + 1) \leq \mathbf{A}(\mathbf{x}, \mathbf{k})$ oraz

$$\forall_{\mathbf{x} \in \mathbb{A}} \lim_{\mathbf{k} \to \infty} \mathbf{A}(\mathbf{x}, \mathbf{k}) = \mathbf{f}(\mathbf{x}). \tag{9}$$

Obliczalne funkcje rzeczywiste

Definicja (obliczalna funkcja rzeczywista)

Funkcja $\mathbf{f} : \mathbb{A} \to \mathbb{R}$ nazywa się *obliczalną*, jeżeli istnieje obliczalna funkcja $\mathbf{A} : \mathbb{A} \times \mathbb{N} \to \mathbb{Q}$, która spełnia

$$\forall_{x \in \mathbb{B}} \, |\mathbf{f}(\mathbf{x}) - \mathbf{A}(\mathbf{x}, \mathbf{k})| < 1/\mathbf{k}. \tag{10}$$

- Funkcja, która jest semiobliczalna od góry i od dołu, jest obliczalna.
- Złożoność Kołmogorowa jest semiobliczalna od góry, ale nie jest obliczalna.

Ciągi i miary

Ciągi nieskończone piszmy jako $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots)$, gdzie $\mathbf{x}_\mathbf{i} \in \mathbb{B}$.
Ponadto oznaczamy $\mathbf{x}_1^\mathbf{m} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\mathbf{m})$.
Symbol $\mathbf{P}$ oznaczać będzie miarę na nieskończonych ciągach.
Będziemy pisać $\mathbf{P}(\mathbf{w}) = \mathbf{P}\left(\left\{\mathbf{x} : \mathbf{x}_1^{|\mathbf{w}|} = \mathbf{w}\right\}\right)$.
Zatem $\mathbf{P}$ spełnia

$$\mathbf{P}(\lambda) = \mathbf{1},$$
$$\mathbf{P}(\mathbf{w}) \geq \mathbf{0},$$
$$\mathbf{P}(\mathbf{w}) = \sum_{\mathbf{a} \in \mathbb{B}} \mathbf{P}(\mathbf{wa}).$$

Miarą obliczalną nazywać będziemy miarę taką, że funkcja $\mathbf{P} : \mathbb{B}^* \ni \mathbf{w} \mapsto \mathbf{P}(\mathbf{w}) \in \mathbb{R}$ jest obliczalna.

Twierdzenie Barrona

Długość dowolnego kodu bezprzedrostkowego jest prawie na pewno asymptotycznie nieskończenie większa niż punktowa entropia.

Twierdzenie Barrona

Niech $\mathbf{B} : \mathbb{B}^* \to \mathbb{B}^*$ będzie kodem bezprzedrostkowym. Wówczas dla dowolnej miary $\mathbf{P}$ mamy

$$\mathbf{P}\left(\left\{\mathbf{x} : \lim_{\mathbf{m}\to\infty} \left[|\mathbf{B}(\mathbf{x}_1^{\mathbf{m}})| + \log \mathbf{P}(\mathbf{x}_1^{\mathbf{m}})\right] = \infty\right\}\right) = \mathbf{1}. \tag{11}$$

Dowód: Nierówność Markowa + nierówność Krafta + lemat Borela-Cantellego.

Ciągi algorytmicznie losowe

Złożoność Kołmogorowa jest najmniejszą (z dokładnością do stałej) semiobliczalną od góry długością kodu bezprzedrostkowego. Mamy stąd:

$$\mathbf{K(x_1^m) \leq -\log P(x_1^m) + 2\log m + c.}$$

Definicja (ciąg algorytmicznie losowy)

Mówimy, że $\mathbf{x}$ jest algorytmicznie losowy, jeżeli

$$\lim_{\mathbf{m}\to\infty} \left[\mathbf{K(x_1^m) + \log P(x_1^m)}\right] = \infty. \tag{12}$$

Z twierdzenia Barrona, zbiór ciągów losowych ma miarę $\mathbf{1}$.

Semimiary

Semimiarą $\mathbf{U}$ nazywamy funkcję, która spełnia

$$\begin{aligned} \mathbf{U}(\lambda) &\leq 1, \\ \mathbf{U}(w) &\geq 0, \\ \mathbf{U}(w) &\geq \sum_{a \in \mathbb{B}} \mathbf{U}(wa). \end{aligned}$$

Twierdzenie

Ciąg $\mathbf{x}$ jest algorytmicznie losowy wtedy i tylko wtedy, gdy

$$\sup_{n \in \mathbb{N}} \frac{\mathbf{U}(x_1^n)}{\mathbf{P}(x_1^n)} < \infty \tag{13}$$

zachodzi dla dowolnej semiobliczalnej od dołu semimiary $\mathbf{U}$.

Zatem ciągi losowe to są dokładnie te ciągi, które są optymalnie kompresowane przez miarę $\mathbf{P}$ w klasie semimiar semiobl. od dołu.

Parametry i jądra

Teraz pokażemy, jak pojęcie losowości algorytmicznej pozwala zrozumieć mocne i słabe strony wnioskowania bayesowskiego.

Niech $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\theta_3}, \ldots)$, gdzie $\boldsymbol{\theta_i} \in \mathbb{B}$, oznacza parametr. Równoważnie ciąg ten będziemy interpretować jako rozwinięcie binarne liczby rzeczywistej, oznaczanej również przez $\boldsymbol{\theta}$.

Definicja (jądro)

Funkcję $\mathbf{P}(\cdot|\cdot) : \mathbb{B}^* \times \mathbb{B}^{\mathbb{N}} \to \mathbb{R}$ nazywamy *jądrem*,
jeżeli $\mathbf{P}(\mathbf{w}|\cdot)$ jest mierzalne dla każdego $\mathbf{w}$,
a $\mathbf{P}(\cdot|\theta)$ jest miarą prawdopodobieństwa dla każdego $\boldsymbol{\theta}$.

Jądra obliczalne

- Jądro $\mathbf{P}$ nazywamy obliczalnym, jeżeli funkcja $\mathbf{P}(\cdot|\cdot) : \mathbb{B}^* \times \mathbb{B}^{\mathbb{N}} \ni (\mathbf{w}, \boldsymbol{\theta}) \mapsto \mathbf{P}(\mathbf{w}|\boldsymbol{\theta}) \in \mathbb{R}$ jest obliczalna. Jest to typowa sytuacja w statystyce.
- Zwróćmy uwagę, że nawet dla obliczalnego jądra, warunkowa miara $\mathbf{P}(\cdot|\boldsymbol{\theta})$ nie musi być obliczalna, jeżeli ustalimy określoną wartość parametru $\boldsymbol{\theta}$.
- Typowo, mając $\mathbf{P}(\cdot|\boldsymbol{\theta})$ możemy obliczyć $\boldsymbol{\theta}$ z dowolną dokładnością, generując typowe dane i estymując z nich parametr. Wówczas miara $\mathbf{P}(\cdot|\boldsymbol{\theta})$ nie jest obliczalna, jeżeli parametr jest algorytmicznie losowy.
- Zatem optymalny obliczalny kompresor danych typowych dla miary warunkowej $\mathbf{P}(\cdot|\boldsymbol{\theta})$ jest różny od tej miary.

Miara bayesowska

- Aby wskazać optymalny kompresor danych typowych dla miary nieobliczalnej, wprowadźmy pewne pojęcia z zakresu wnioskowania bayesowskiego.
- Niech $\mathbf{Q}$ będzie obliczalną miarą zwaną rozkładem a priori, zaś $\mathbf{P}$ niech będzie obliczalnym jądrem. Miara

$$\mathbf{Y}(\mathbf{w}) = \int \mathbf{P}(\mathbf{w}|\theta)\mathbf{dQ}(\theta) \tag{14}$$

  jest obliczalna i będzie nazywana miarą bayesowską.
- Pokażemy, że miara bayesowska $\mathbf{Y}$ jest optymalnym obliczalnym kompresorem danych typowych dla miary warunkowej $\mathbf{P}(\cdot|\theta)$, jeżeli parametr jest typowy dla rozkładu a priori $\mathbf{Q}$.

Bayes jest optymalny prawie wszędzie

Oznaczmy zbiór ciągów algorytmicznie losowych dla miary **Y** jako

$$\mathcal{R}_{\mathbf{Y}} := \left\{ \mathbf{x} : \lim_{\mathbf{m}\to\infty} \left[\mathbf{K}(\mathbf{x}_1^{\mathbf{m}}) + \log \mathbf{Y}(\mathbf{x}_1^{\mathbf{m}}) \right] = \infty \right\}. \tag{15}$$

Następujące twierdzenie orzeka, że dla prawie wszystkich parametrów $\boldsymbol{\theta}$, dane typowe dla miar warunkowych $\mathbf{P}(\cdot|\boldsymbol{\theta})$ są optymalnie kompresowane przez miarę **Y**.

Twierdzenie

Mamy

$$\mathbf{Q}\left(\{\boldsymbol{\theta} : \mathbf{P}(\mathcal{R}_{\mathbf{Y}}|\boldsymbol{\theta}) = 1\}\right) = 1.$$

Dowód: Prosty wniosek z $\mathbf{Y}(\mathcal{R}_{\mathbf{Y}}) = 1$.

Bayes jest optymalny tylko dla losowych parametrów

W następnej kolejności, chcielibyśmy pokazać, że dla pewnych jąder poprzednie twierdzenie może zostać wzmocnione jako

$$\mathbf{P}(\mathcal{R}_{\mathbf{Y}}|\theta) = \begin{cases} \mathbf{1} & \text{jeżeli } \theta \in \mathcal{R}_{\mathbf{Q}}, \\ \mathbf{0} & \text{jeżeli } \theta \notin \mathcal{R}_{\mathbf{Q}}. \end{cases} \tag{16}$$

Istnieją jądra, dla których nie jest to prawdą.

Na przykład, jeżeli $\mathbf{P}(\cdot|\theta)$ nie zależy od θ, otrzymujemy $\mathbf{P}(\mathcal{R}_{\mathbf{Y}}|\theta) = \mathbf{1}$ dla każdego θ niezależnie od rozkładu a priori.

Pokażemy, że dychotomia (16) zachodzi, jeżeli jądro dopuszcza istnienie efektywnego estymatora.

Efektywny estymator

Definicja (efektywnie ściśle zgodny estymator)

Estymator $\mathbf{T(x_1^n)}$ nazywamy *efektywnie ściśle zgodnym*, jeżeli istnieje obliczalna funkcja $\mathbf{N}(\epsilon, \delta)$ taka, że dla każdego θ i dla wszystkich ϵ i δ mamy

$$\mathbf{P}\left(\left\{\mathbf{x} : \sup_{\mathbf{n} \geq \mathbf{N}(\epsilon,\delta)} \left|\mathbf{T}(\mathbf{x}_1^{\mathbf{n}}) - \theta\right| > \epsilon\right\} \middle| \theta\right) \leq \delta.$$

Ciągi warunkowo losowe

Definicja (ciąg warunkowo losowy)

Mówimy, że ciąg $\mathbf{x}$ jest *warunkowo algorytmicznie losowy* dla obliczalnego jądra $\mathbf{P}$ względem parametru $\boldsymbol{\theta}$, jeżeli

$$\lim_{m\to\infty} \left[K(x_1^m|\theta) + \log P(x_1^m|\theta)\right] = \infty. \tag{17}$$

Zbiór ciągów warunkowo algorytmicznie losowych oznaczamy jako

$$\mathcal{R}_{P|\theta} := \left\{x : \lim_{m\to\infty} \left[K(x_1^m|\theta) + \log P(x_1^m|\theta)\right] = \infty\right\}. \tag{18}$$

Z twierdzenia Barrona, $P(\mathcal{R}_{P|\theta}|\theta) = 1$.

Twierdzenie V'yugina

Twierdzenie V'yugina

Jeżeli jądro $\mathbf{P}$ dopuszcza efektywnie ściśle zgodny estymator $\mathbf{T}(\mathbf{x}_1^n)$, to

$$\lim_{\mathbf{n}\to\infty} \mathbf{T}(\mathbf{x}_1^n) = \theta \tag{19}$$

dla każdego $\mathbf{x} \in \mathcal{R}_{\mathbf{P}|\theta}$.

Wówczas zbiory $\mathcal{R}_{\mathbf{P}|\theta}$ są rozłączne.

Twierdzenie Takahashiego

- Kolejny fakt, z którego skorzystamy, to rozkład zbioru ciągów losowych względem miary bayesowskiej.
- Zbiór ten rozkłada się na zbiory ciągów losowych względem miar warunkowych.
- Należy podkreślić, że rozkład ten przebiega wyłącznie po parametrach, które są losowe względem rozkładu a priori.

Twierdzenie Takahashiego

Dla $\mathbf{Y} = \int \mathbf{P}(\cdot|\theta)\mathbf{d}\mathbf{Q}(\theta)$ zachodzi

$$\mathcal{R}_{\mathbf{Y}} = \bigcup_{\theta \in \mathcal{R}_{\mathbf{Q}}} \mathcal{R}_{\mathbf{P}|\theta}. \tag{20}$$

Twierdzenie o optymalności

Teraz możemy sformułować twierdzenie o optymalności.

Twierdzenie

Dla obliczalnego jądra $\mathbf{P}$ i obliczalnego rozkładu a priori $\mathbf{Q}$ określmy miarę bayesowską $\mathbf{Y} = \int \mathbf{P}(\cdot|\theta)\mathbf{d}\mathbf{Q}(\theta)$. Jeżeli jądro dopuszcza efektywnie ściśle zgodny estymator parametru, to mamy dychotomię

$$\mathbf{P}(\mathcal{R}_{\mathbf{Y}}|\theta) = \begin{cases} \mathbf{1} & \text{jeżeli } \theta \in \mathcal{R}_{\mathbf{Q}}, \\ \mathbf{0} & \text{jeżeli } \theta \notin \mathcal{R}_{\mathbf{Q}}. \end{cases} \tag{21}$$

Dowód.

Z twierdzenia V'yugina zbiory $\mathcal{R}_{\mathbf{P}|\theta}$ są rozłączne. Mamy także $\mathbf{P}(\mathcal{R}_{\mathbf{P}|\theta}|\theta) = \mathbf{1}$ dla każdego θ. Zatem (21) wynika z twierdzenia Takahashiego. □

Modele τ-wyuczalne

Definicja (modele τ-wyuczalne)

Niech $\mathbf{P}$ będzie jądrem, zaś $\mathbf{Q}$ rozkładem a priori. Zdefiniujmy miarę

$$\mathbf{T}(\mathbf{w}, \mathbf{z}) := \int_{\mathbf{A}(\mathbf{z})} \mathbf{P}_\theta(\mathbf{w}) d\mathbf{Q}(\theta), \tag{22}$$

gdzie $\mathbf{A}(\mathbf{z}) := \{\theta \in \Theta : \mathbf{z} \text{ jest przedrostkiem } \theta\}$, oraz miarę

$$\mathbf{Y}(\mathbf{w}) := \mathbf{T}(\mathbf{w}, \lambda) = \int \mathbf{P}_\theta(\mathbf{w}) d\mathbf{Q}(\theta). \tag{23}$$

Parę $(\mathbf{P}, \mathbf{Q})$ nazywam τ-*wyuczalną*, jeżeli dla funkcji $\tau : \mathbb{B}^* \to \mathbb{N}$, wszystkich $\theta \in \mathbb{B}^{\mathbb{N}}$, $\mathbf{P}_\theta$-prawie wszystkich $\mathbf{x}$ i $\mathbf{n} = \tau(\theta_1^{\mathbf{m}})$ zachodzi

$$\lim_{\mathbf{m} \to \infty} \mathbf{T}(\mathbf{x}_1^{\mathbf{n}}, \theta_1^{\mathbf{m}}) / \mathbf{Y}(\mathbf{x}_1^{\mathbf{n}}) = 1. \tag{24}$$

Mój wynik

Wykazałem następujący fakt:

Twierdzenie

Niech para $(\mathbf{P}, \mathbf{Q})$ będzie τ-wyuczalna z obliczalnym jądrem $\mathbf{P}$, obliczalnym rozkładem a priori $\mathbf{Q}$ i obliczalną funkcją τ. Wówczas zachodzi dychotomia

$$\mathbf{P}(\mathcal{R}_{\mathbf{Y}}|\theta) = \begin{cases} \mathbf{1} & \text{jeżeli } \theta \in \mathcal{R}_{\mathbf{Q}}, \\ \mathbf{0} & \text{jeżeli } \theta \notin \mathcal{R}_{\mathbf{Q}}. \end{cases} \tag{25}$$

Dowód tego faktu nie odwołuje się do rozłączności zbiorów $\mathcal{R}_{\mathbf{P}|\theta}$, której nie udało mi się wykazać.

Efektywny estymator II

Definicja (efektywnie ściśle zgodny estymator II)

Estymator $\mathbf{T(x_1^n)}$ nazywamy *efektywnie ściśle zgodnym II*, jeżeli dla każdego θ istnieje obliczalna funkcja $\mathbf{N}(\epsilon, \delta)$ taka, że dla wszystkich ϵ i δ mamy

$$\mathbf{P}\left(\left\{\mathbf{x} : \sup_{\mathbf{n} \geq \mathbf{N}(\epsilon,\delta)} \left|\mathbf{T}(\mathbf{x}_1^n) - \theta\right| > \epsilon\right\}\middle|\theta\right) \leq \delta.$$

Czy tw. V'yugina można uogólnić na ten przypadek?

Podsumowanie

- Pokazaliśmy, że jeżeli parametr można efektywnie wyestymować, to miara bayesowska zadaje optymalną obliczalną kompresję danych, które są losowe względem miary warunkowej, wtedy i tylko wtedy, gdy parametr jest losowy względem rozkładu a priori.
- Stwierdzenie to jest użyteczne, jeżeli miara warunkowa nie jest obliczalna dla zadanego parametru.
- Ponadto, wiedząc, kiedy kompresja bayesowska jest suboptymalna, powinniśmy dostosowywać rozkład a priori do naszych rzeczywistych przypuszczeń co do złożoności algorytmicznej parametru.
- Czy dla dowolnego parametru istnieje obliczalny rozkład a priori, względem którego parametr jest losowy?