RZECZ O ROZKŁADACH A PRIORI

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

Plan wykładu

1) Rozkład *a priori* we wnioskowaniu bayesowskim
2) Sprzężone rozkłady *a priori*
3) Nieinformacyjne rozkłady *a priori*
4) Niewłaściwe rozkłady *a priori*
5) Podsumowanie
6) Przykłady

Rozkład *a priori* we wnioskowaniu bayesowskim 1

- Przypomnijmy wzór Bayesa:

$$p(\theta|y) = \frac{p(y,\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}$$

- Statystyka bayesowska „oferuje" w pełni probabilistyczny – bo za pomocą „całego" rozkładu prawdopodobieństwa, $p(\theta|y)$ – obraz niepewności związanej z wnioskowaniem o parametrach (zamiast tylko oceny punktowej, błędu szacunku, czy nawet przedziałów ufności). Jednak formalnie poprawne – bo za pomocą formalnie prawidłowej formuły, tj. wzoru Bayesa – uzyskanie tego rozkładu (czyli rozkładu *a posteriori*) wymaga obecności rozkładu *a priori*, $p(\theta)$, który to rozkład musimy zadać „w jakiś sposób", choćbyśmy nawet nie chcieli „zanieczyścić" przezeń tej informacji o parametrach, którą niosą ze sobą dane (a która „wchodzi" do modelu poprzez rozkład próbkowy, $p(y|\theta)$, czyli funkcję wiarygodności)
- Skoro musimy tu (czyli na gruncie statystyki bayesowskiej) wprowadzić jakąś (nie)wiedzę o parametrach *a priori* do modelu statystycznego (czego nie ma na gruncie „klasycznym"), to można zadać pytanie: *Czy to dobrze, czy niedobrze*??

Rozkład *a priori* we wnioskowaniu bayesowskim 2

- Różne aspekty obecności rozkładu *a priori* w modelowaniu bayesowskim:
  - Zalety:
    1) umożliwia wprowadzenie do modelu statystycznego dodatkowej (poza samymi tylko danymi) informacji/wiedzy o parametrach = „naturze” modelowanego zjawiska, w tym wiedzy eksperckiej czy też wynikającej ze znajomości podobnych badań
    2) ta z poprzedniej strony: umożliwia skorzystanie ze wzoru Bayesa, w wyniku którego uzyskujemy cały rozkład prawdopodobieństwa dla parametrów
  - Wady/ograniczenia:
    1) Sam fakt konieczności wprowadzenia informacji *a priori* do modelu (a my może byśmy właśnie tego nie chcieli?)

       → ALE: rozkłady nieinformacyjne (w dalszej części tego materiału) lub w przybliżeniu takowe
    2) Stwarzają możliwość manipulacji („forsowania”) wyników *a posteriori* – poprzez odpowiednie zacieśnienie rozkładu *a priori* nie pozwalamy danym „przemówić” zbyt swobodnie

       → ALE:

       - taki dogmatyzm jest brakiem uczciwości badawczej :/

       - możemy przeprowadzić **analizę wrażliwości** rozkładu *a posteriori* na różną informację *a priori*

       - wpływ rozkładu *a priori* na *a posteriori* maleje wraz ze wzrostem liczebności próby

Sprzężone rozkłady *a priori* 1

- **Definicja:**

  Mówimy, że rozkład *a priori* danego typu (tj. należący do danej rodziny rozkładów prawdopodobieństwa, np. gamma, beta, normalny) jest **rozkładem sprzężonym** (ang. *conjugate prior*) – w domyśle: sprzężonym z daną funkcją wiarygodności (rozkładem próbkowym) – wtedy i tylko wtedy, gdy uzyskiwany przy nim **rozkład *a posteriori* jest tego samego typu, co ów rozkład *a priori*** (tj. należy do tej samej rodziny rozkładów p-stwa, np. – odpowiednio – gamma, beta, normalny)

- **Przykłady:**
  - Model dla obserwacji z rozkładu wykładniczego (tj. $y_t|\lambda \sim iiExp(\lambda)$, $t = 1,2, \dots, T$):

    $$p(\lambda) = f_G(\lambda|a, b) \to p(\lambda|y) = f_G(\lambda|\bar{a}, \bar{b})$$

  - Model dla obserwacji z rozkładu Poissona (tj. $y_t|\lambda \sim iiPois(\lambda)$, $t = 1,2, \dots, T$):

    $$p(\lambda) = f_G(\lambda|a, b) \to p(\lambda|y) = f_G(\lambda|\bar{a}, \bar{b})$$

  - Model dla prób Bernoulliego (tj. $y_t|\varphi \sim iiDwupunkt.(\varphi)$, $t = 1,2, \dots, T$):

    => $p(y_t|\varphi) = \varphi^{y_t}(1-\varphi)^{1-y_t}, \quad y_t \in \{0, 1\}, \quad \varphi \in [0, 1]$:

    $$p(\varphi) = f_{Be}(\varphi|a, b) \to p(\varphi|y) = f_{Be}(\varphi|\bar{a}, \bar{b})$$

- **Pełny wykaz rozkładów sprzężonych dla prostych modeli jednoparametrycznych** – następna strona

Sprzężone rozkłady *a priori* 2

- **Pełny wykaz rozkładów sprzężonych dla prostych modeli jednoparametrycznych**

Typ rozkładu obserwacji	Rozkład próbkowy pojedynczej obserwacji, $p(y_t \mid \theta)$	Sprzężony rozkład *a priori*	Uwagi
Dyskretny	Dwupunktowy (ang. *Bernoulli distribution*)	Beta	
	Dwumianowy (ang. *binomial distribution*; pl. rozkład Bernoulliego)	Beta	
	Pascala (tzw. ujemny rozkład dwumianowy)	Beta	
	Geometryczny	Beta	
	Wielomianowy	Dirichlet	
	Poissona	Gamma	
Ciągły	Gamma – przy ustalonej wartości parametru kształtu (tj. tego pierwszego)	Gamma – tylko dla parametru „rate", tj. b_y w rozkładzie $y_t \mid a_y, b_y \sim G(a_y, b_y)$, $E(y_t \mid a_y, b_y) = \dfrac{a_y}{b_y}$	Dla parametru kształtu, tj. a_y nie istnieje sprzężony rozkład *a priori*. Rozkład gamma dla b_y jest *warunkowo* sprzężonym rozkładem *a priori* – bo przy ustalonej wartości a_y
	Wykładniczy	Gamma	Rozkład wykładniczy = szczególny przypadek rozkładu gamma (z parametrem kształtu równym 1)
	Normalny, $N(\mu, \sigma^2)$ – dla μ, przy ustalonej σ^2	Normalny	
	Normalny, $N(\mu, \sigma^2)$ – dla σ^2, przy ustalonej μ	Odwrócony gamma – dla σ^2 (wariancja) Gamma – dla $\tau \equiv (\sigma^2)^{-1}$ (precyzja)	

- Często wykorzystywane **w praktyce**! But why??!
- OK, ale jak zadawać wartości hiperparametrów? → To wprowadzi „jakąś" informację *a priori*

Nieinformacyjne rozkłady *a priori 1*

- Zwane też obiektywnymi (ang. *objective priors*) albo referencyjnymi (ang. *reference priors*)
- Choć w zamyśle ma wyrażać mój brak wstępnej wiedzy o parametrach, w istocie nie istnieje coś takiego jak *dosłownie* **nie**informacyjny rozkład *a priori* – …
- Istnieje wiele koncepcji wyznaczania rozkładów nieinformacyjnych – w czasie naszych zajęć poznamy tylko jedną z nich: **regułę Jeffreysa**, opartą na macierzy informacyjnej Fishera :) (o czym za chwilę)
- Kiedy sięgamy po rozkłady nieinformacyjne, a kiedy informacyjne (czyli nie zależy nam aż „tak bardzo" na nie-informacyjności):
    - Rozkłady informacyjne – w zagadnieniach związanych z podejmowaniem decyzji (np. biznesowych), gdy chcemy „wzmocnić" informację, którą niosą dane, o naszą wiedzę ekspercką, doświadczenie, czy wyniki wcześniejszych, analogicznych badań
    - Rozkłady nieinformacyjne – w zagadnieniach wnioskowania naukowego, gdzie – jak w laboratorium – zależy nam na pewnej „sterylności", aby naszą wiedzą nie „zanieczyścić" informacji płynącej z danych (albo raczej: „zanieczyścić" w możliwie najmniejszym stopniu)

→ Problemy:

- Samo wyznaczenie rozkładu nieinformacyjnego (dla danego rozkładu próbkowego) – da się go wyznaczyć tylko w relatywnie prostych modelach (większości jednoparametrycznych i KMNRL, ale już nie dla modeli GARCH czy SV)
- Zwykle (choć z nielicznymi wyjątkami) wyprowadzony rozkład nieinformacyjny okazuje się tzw. *niewłaściwym* rozkładem prawdopodobieństwa (o czym później), co komplikuje nie tyle same wyznaczenie rozkładu *a posteriori* (choć też może), co następnie porównywanie różnych, konkurencyjnych modeli bayesowskich do opisu tego samego zjawiska (na gruncie bayesowskim nie „załatwiamy" tego za pomocą R^2 czy nawet kryteriów informacyjnych…)

Nieinformacyjne rozkłady *a priori 2*

- **Reguła Jeffreysa:**

Niech $l_y(\theta) \equiv \ln L_y(\theta)$, zaś $I(\theta)$ oznacza macierz informacyjną Fishera:

$$I(\theta) = -E_{y|\theta}\left(\frac{\partial^2 l_y(\theta)}{\partial\theta\partial\theta'}\right)$$

Jeżeli $p(\theta) \propto \sqrt{\det I(\theta)}$, to rozkład $p(\theta)$ wprowadza do modelu najmniejszą (w porównaniu do funkcji wiarygodności) informację o θ. Taki rozkład będziemy określać mianem nieinformacyjnego rozkładu Jeffreysa i oznaczać symbolem $p_J(\theta)$.

Niewłaściwe rozkłady *a priori* 1

- **Definicja:** <u>Niewłaściwym rozkładem *a priori*</u> (wektorowego) parametru $\theta = (\theta_1\ \theta_2\ \ldots\ \theta_s)' \in \Theta \subseteq \mathbb{R}^s$ nazywamy „rozkład" prawdopodobieństwa o takiej „funkcji gęstości" $p(\theta)$, że
$$\int_\Theta p(\theta)d\theta = +\infty,$$
ale Θ da się rozbić na sumę przeliczalnie wielu rozłącznych podzbiorów miary skończonej.

- **Przykłady** (wszystkie one okażą się być nieinformacyjne w modelach regresji liniowej)

- $p(\theta) \propto c,\quad c > 0$ – dowolna stała dodatnia, zaś $\theta \in \mathbb{R}$ (tzw. *flat prior*) → r. jednostajny na $\mathbb{R}$
- $p(\theta) \propto c,\quad c > 0$ – dowolna stała dodatnia, zaś $\theta \in \mathbb{R}^2$ (tzw. *flat prior*) → r. jednostajny na $\mathbb{R}^2$
- $p(\theta) \propto \theta^{-1},\quad \theta \in \mathbb{R}_+$ → ramię hiperboli w I ćwiartce: $\int_0^{+\infty} \theta^{-1} d\omega = \cdots = +\infty$

- Po niewłaściwe rozkłady *a priori* nie sięgamy z jakiejkolwiek innej przyczyny, jak tylko wtedy gdy okazują się po prostu rozkładami nieinformacyjnymi

- Okazuje się, że bardzo często (choć NIE zawsze) przy niewłaściwym rozkładzie *a priori*, uzyskiwany rozkład *a posteriori* jest jednak rozkładem właściwym, przez co da się przeprowadzić jego analizę
  - → Wtedy jednak znacząco komplikuje się porównywanie konkurencyjnych modeli (opowiemy sobie o tym w przyszłości)

Niewłaściwe rozkłady *a priori* 2

- Powyższe przykłady okażą się nieinformacyjne w modelach normalnej regresji liniowej:

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t = x_t\beta + \varepsilon_t, \quad \varepsilon_t \sim iiN(0, \sigma^2 = \tau^{-1}), \quad \tau = Prec(\varepsilon_t)$$

  - $p(\beta_i) \propto c > 0, \qquad p(\beta) \propto c > 0$ (*flat priors*)
  - $p(\tau) \propto \tau^{-1}$

→ Zamiast $p(\beta_i) \propto c$ można by przyjąć rozkład „w przybliżeniu" płaski, np. normalny o dużej wariancji (ang. *vague priors, diffuse priors*) = rozkład w przybliżeniu nieinformacyjny

→ Dla $p(\tau) \propto \tau^{-1}$ można zauważyć, że $\tau^{-1} = \tau^{0-1} \underbrace{e^{-0\cdot\tau}}_{1} \propto f_G(\tau|a \to 0, b \to 0)$, a zatem rozkładem w przybliżeniu nieinformacyjnym dla precyzji składników losowych byłby tu rozkład gamma o hiperparametrach a i b bliskich 0

- UWAGA ogólna: Gdyby przy danym niewłaściwym rozkładzie *a priori* rozkład *a posteriori* okazał się także niewłaściwy (czyli $\int_\Theta p(\theta|y)d\theta = +\infty$), wówczas mówimy, że rozkład *a posteriori* <u>nie istnieje</u>

Podsumowanie

- **Zalety stosowania rozkładów Jeffreysa** (jeśli da się go w ogóle wyznaczyć dla danej postaci funkcji wiarygodności, bo nie zawsze – tj. nie dla wszystkich modeli próbkowych – jest to możliwe):
  - nieinformacyjność
  - niezmienniczość względem wzajemnie jednoznacznych reparametryzacji modelu (tj. jeśli mamy rozkład Jeffreysa np. dla λ, a chcemy zreparametryzować model przyjmując $\omega = \frac{1}{\lambda}$, to jeśli przekształcimy $p_J(\lambda)$ odpowiednio na rozkład dla ω, to ów rozkład dla ω także będzie nieinformacyjny (w sensie reguły Jeffreysa))

- **Wady/ograniczenia stosowania rozkładów Jeffreysa** – te wskazane już wcześniej:
  - Problem może stanowić już samo wyznaczenie rozkładu nieinformacyjnego (dla danego rozkładu próbkowego) – jest to możliwe tylko w relatywnie prostych modelach (większości jednoparametrycznych i KMNRL, ale już nie dla modeli GARCH czy SV)
  - Zwykle (choć z nielicznymi wyjątkami) wyprowadzony rozkład nieinformacyjny okazuje się tzw. ***niewłaściwym*** **rozkładem prawdopodobieństwa**, co komplikuje nie tyle same wyznaczenie rozkładu *a posteriori* (choć też może), co następnie porównywanie różnych, konkurencyjnych modeli bayesowskich do opisu tego samego zjawiska

  → ALE:

- Można pokazać, że te rozkłady Jeffreysa, które są niewłaściwe (znacząca większość), stanowią <u>graniczne przypadki rozkładów sprzężonych</u> (przy hiperparametrach dążących np. do 0), co można wykorzystać do zadania rozkładu *a priori* **w przybliżeniu** nieinformacyjnego (odpowiednio zadając wartość hiperparametru bliską właśnie tej jego granicznej wartości, np. 0)

Podsumowanie

- **Co należy każdorazowo rozeznać przy specyfikowaniu rozkładu *a priori*?**

1) Czy dla danego rozkładu próbkowego istnieje *sprzężony* rozkład *a priori*?

→ wygoda/pragmatyzm sięgania po sprzężone rozkłady *a priori*

→ zauważmy(!): rozkłady sprzężone są zawsze właściwymi rozkładami p-stwa

→ z reguły są „mniej lub bardziej” informacyjne (musimy jakoś „arbitralnie” zadać wartości hiperparametrów, zwykle zapewniając odpowiednie rozproszenie), choć czasem okazuje się, że dla pewnych wartości hiperparametrów uzyskujemy właściwy rozkład nieinformacyjny Jeffreysa (jak w przypadku modelu dla prób Bernoulliego)

Podsumowanie

➢ **Co należy każdorazowo rozeznać przy specyfikowaniu rozkładu *a priori*?**

2) <u>Czy dla danego rozkładu próbkowego istnieje *nieinformacyjny* rozkład *a priori*</u>?

→ pożądane we wnioskowaniu naukowym (w odróżnieniu od kontekstu decyzyjnego)

→ czy wyznaczony (wg reguły Jeffreysa) rozkład jest *właściwym* rozkładem p-stwa?

TAK: czy jest sprzężony?

NIE: czy rozkład *a posteriori* jest wówczas *właściwym* rozkładem p-stwa?

TAK: zatem rozkład *a posteriori* istnieje i da się go analizować, ALE gdybyśmy chcieli porównać moc wyjaśniającą takiego modelu (z niewłaściwym rozkładem *a priori*) z innymi, konkurencyjnymi modelami (np. o innych rozkładach *a priori*), to będzie to co najmniej problematyczne… Lepiej wtedy zastąpić niewłaściwy rozkład *a priori* jakimś rozkładem właściwym (najlepiej sprzężonym) o możliwie podobnym kształcie (np. bardzo rozproszony rozkład normalny jako „proteza" = przybliżenie rozkładu jednostajnego na całej osi liczb rzeczywistych)

NIE: zatem rozkład *a posteriori* NIE istnieje, więc jego analiza jest po prostu niemożliwa

Przykłady

Zadanie 1.1 [c.d. Zadania 1: model dla obserwacji z rozkładu wykładniczego]

Wyznacz nieinformacyjny rozkład *a priori* dla parametru λ.

Zadanie 2.1 [c.d. Zadania 2: model dla obserwacji z rozkładu Poissona]

Wyznacz nieinformacyjny rozkład *a priori* dla parametru λ.

Zadanie 3.1 [c.d. Zadania 3: model dla prób Bernoulliego]

Jak wyglądałby Twoim zdaniem, tak intuicyjnie, nieinformacyjny rozkład *a priori* dla odsetka studentów pracujących, φ? Wyznacz nieinformacyjny rozkład *a priori* stosując regułę Jeffreysa i porównaj z rozkładem „intuicyjnie nieinformacyjnym”

- **W każdym z powyższych zadań sprawdź:**
  - Czy wyznaczony nieinformacyjny rozkład *a priori* jest rozkładem sprzężonym?
  - Czy wyznaczony nieinformacyjny rozkład *a priori* jest właściwym rozkładem p-stwa?
    - Jeśli nie, to czy rozkład uzyskiwany *a posteriori* jest właściwy?