*Wnioskowanie bayesowskie w ekonomii empirycznej (*Analityka gospodarcza, zima 2023/2024)

***Bayesowska analiza Modelu Normalnej Regresji Liniowej***

***/Bayesowski Model Normalnej Regresji Liniowej,* BMNRL */***

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

**Plan wykładu**

1. Konstrukcja BMNRL
   1. Rozkład próbkowy (funkcja wiarygodności) – wynika z KMNRL
   2. Rozkład(y) *a priori*
2. Charakterystyka rozkładu(-ów) *a posteriori*
3. Testowanie pojedynczego parametru strukturalnego (w tym jego istotności)

**Konstrukcja BMNRL – 1**

* Zgrubne przypomnienie
  + Równanie modelu regresji liniowej:
    - Zapis skalarny – dla każdego (= liczba modelowanych obserwacji):
    - Zapis macierzowy:

* + Składniki losowe:
    - Zapis skalarny (): , ,
    - Zapis macierzowy: ,

gdzie , (Prec = 1/Var, ale tylko w r. normalnym!)

**Konstrukcja BMNRL – 2**

* Wektor parametrów modelu:
* Model bayesowski:
  + Rozkład próbkowy (funkcja wiarygodności): (wynika z założeń KMNRL)
  + Rozkład *a priori*:

**Konstrukcja BMNRL – rozkład próbkowy**

* Założenia KMNRL – w zapisie macierzowym

1. – znana macierz nielosowa
2. () ()

Założenia 4-6 ⇔

* Pytanie o :
  + Skoro i , to ,

gdzie ,

* + Zatem

**Konstrukcja BMNRL – rozkład(y) *a priori***

**Konstrukcja BMNRL – rozkład(y) *a priori* 1**

* Cztery typowe specyfikacje :
  + **[RJz]** **Reguła Jeffreysa (rozkład nieinformacyjny) – z zależnością** pomiędzy i
  + **[RJn]** **Reguła Jeffreysa (rozkład nieinformacyjny) – z niezależnością** pomiędzy i
  + **[GNz]** Rozkład **gamma-normalny – z zależnością** pomiędzy i
  + **[GNn]** Rozkład **gamma-normalny – z niezależnością** pomiędzy i

**Konstrukcja BMNRL – rozkład(y) *a priori* 2**

1. **[RJz]** **Reguła Jeffreysa – z zależnością** pomiędzy i

i ⇒

🡪 W jaki sposób zależy od ?

|  |  |
| --- | --- |
|  | Uwaga:  🡨 to nie są wykresy funkcji *gęstości*, tylko wyrażenia , którego nie da się unormować |

🡪 Wady i zalety:

+ rozkład nieinformacyjny

+ rozkład *a posteriori* należy do znanej rodziny rozkładów p-stwa ⇒ wzory analityczne dla r. *a posteriori*

− rozkład niewłaściwy, ale r. *a posteriori* jest właściwy (niemniej, niewłaściwy r. *a priori* utrudnia kwestię porównywania konkurencyjnych modeli – a tu będzie to ważne…)

**Konstrukcja BMNRL – rozkład(y) *a priori* 3**

1. **[RJn]** **Reguła Jeffreysa – z niezależnością** pomiędzy i

i

(*flat prior*) i

⇒

🡪 Wady i zalety (jak dla RJz):

+ rozkład nieinformacyjny

+ rozkład *a posteriori* należy do znanej rodziny rozkładów p-stwa ⇒ wzory analityczne dla r. a posteriori

− rozkład niewłaściwy, ale r. *a posteriori* jest właściwy (niemniej, niewłaściwy r. *a priori* utrudnia kwestię porównywania konkurencyjnych modeli – a tu będzie to ważne…)

🡪 Ciekawostka: przy tym rozkładzie *a priori*, wnioskowanie bayesowskie o parametrach strukturalnych „pokrywa się” z podejściem niebayesowskim = bayesowska reinterpretacja wyników „klasycznej” estymacji

* 🡨 Metoda Najmniejszych Kwadratów, MNK
* Przedziały *HPD* dla parametrów strukturalnych pokrywają się z „klasycznymi” przedziałami ufności (lecz mają inną – zdecydowanie bardziej naturalną – interpretację)

**Konstrukcja BMNRL – rozkład(y) *a priori* 4**

1. **[GNz]** Rozkład **gamma-normalny – z zależnością** pomiędzy i (hiperparametry)

i i

🡪 Taki rozkład jest bardzo „wygodny”/pragmatyczny, albowiem:

⇒ 🡨 **-wymiarowy rozkład *t*-Studenta**

(a rozkład *t*-Studenta jest dobrze „znany i lubiany”, więc łatwo z nim pracować)

– liczba stopni swobody; – wektor niecentralności (także , gdy )

🡪 – macierz precyzji

⇒ , gdy Uwaga:

⇒ Jeśli zadajemy , (alternatywnie ) oraz , to hiperparametry:

, ,

⇒ 🡨 **1-wymiarowy rozkład *t*-Studenta**

⇒ ⇒

Uwaga:

**Konstrukcja BMNRL – rozkład(y) *a priori* 5**

1. **[GNz]** Rozkład **gamma-normalny – z zależnością** pomiędzy i – **C.D.**

🡪 Wady i zalety:

+ rozkład właściwy

+ rozkład sprzężony ⇒ rozkład *a posteriori* jest tego samego typu (tj. gamma-normalny), więc należy do znanej rodziny rozkładów p-stwa ⇒ wzory analityczne dla r. *a posteriori*

+ może być wykorzystany do specyfikacji „w przybliżeniu” nieinformacyjnego rozkładu *a priori* – *a’la* [RJ**z**]:

przy (*vague prior, diffuse prior*)

przy i

− rozkład informacyjny (ALE i tak wraz z rola rozkładu *a priori* maleje)

− określenie wstępnej wiedzy o zależności pomiędzy i (poprzez ) może być nieco problematyczne, ale zwykle przyjmuje się, że jest pewną macierzą diagonalną (zera poza przekątną)

**Konstrukcja BMNRL – rozkład(y) *a priori* 6**

1. **[GNn]** Rozkład **gamma-normalny – z niezależnością** pomiędzy i

i i

⇒

⇒ Jeśli zadajemy , (alternatywnie ), to hiperparametry (jak w GNz):

,

🡪 Wady i zalety:

+ rozkład właściwy

+ nie ma potrzeby określania wstępnej wiedzy o zależności pomiędzy i

+ może być wykorzystany do specyfikacji „w przybliżeniu” nieinformacyjnego rozkładu *a priori* – *a’la* [RJ**n**]:

przy (*vague prior, diffuse prior*)

przy i

− rozkład informacyjny, ALE wraz z rola rozkładu *a priori* maleje

− nie jest rozkładem sprzężonym i, co więcej, rozkład *a posteriori* NIE należy do żadnej znanej rodziny rozkładów p-stwa ⇒ BRAK wzorów analitycznych na charakterystyki r. *a posteriori* ⇒ konieczność sięgania po zaawansowane metody symulacyjne w celu uzyskania próby (pseudo-)losowej z r. *a posteriori* – tzw. metody Monte Carlo oparte na łańcuchach Markowa; ang. *Markov Chain Monte Carlo*, **MCMC**

**Charakterystyka rozkładu *a posteriori –* rozkład łączny**

* Poniżej omówimy **tylko przypadki II (RJn) i III (GNz),** gdyż:
* I (RJz) – rzadko wykorzystywany w praktyce
* IV (GNn) – brak formuł analitycznych 🡪 konieczność stosowania zaawansowanych metod numerycznych (MCMC)
* Rozkład *a posteriori* – **zarówno w II (RJn), jak i III (GNz)** – jest rozkładem **gamma-normalnym**:

|  |  |
| --- | --- |
| **II (RJn)** | **III (GNz)** |
|  |  |
|  |  |
|  |  |
|  |  |

🡪 Najbardziej nurtujące nas PYTANIE:

**Charakterystyka rozkładu *a posteriori* – rozkłady brzegowe 1**

* **Brzegowy rozkład *a posteriori* dla**  – już mamy:
* Pozostałe charakterystyki – numerycznie (w pakietach lub symulacyjnie)
* **Brzegowy rozkład *a posteriori* dla**  (czyli łączny dla wszystkich parametrów strukturalnych)

,

gdzie

* , dla
* ( – dowolne)
* mediana w przypadku rozkładów wielowymiarowych – za trudne…)
* , dla →

🡪 Zauważmy, że przy :

**Charakterystyka rozkładu *a posteriori* – rozkłady brzegowe 2**

* **Brzegowe rozkłady *a posteriori* dla pojedynczych**  **():**

gdzie jest -tą współrzędną wektora , natomiast

Uwaga:

* dla
* ( – dowolne)
* dla →
* , – dwustronna wartość kryt. w
  + , gdzie ;
  + w Excelu: → =ROZKŁAD.T.ODW(*prawdopodobieństwo* = ; *stopnie\_swobody* = ) (starsza wersja)

→ =ROZKŁ.T.ODWR.DS(*prawdopodobieństwo* = ; *stopnie\_swobody* = ) (nowsza wersja)

**Charakterystyka rozkładu *a posteriori* – [RJn] = bayesowska reintepretacja MNK**

* **Zauważmy**, że w przypadku **[RJn]** otrzymujemy **konkretnie**:
  + , , dla
  + , ( – dowolne)
  + , dla

🡪Zauważmy, że , czyli bayesowska ocena niepewności wnioskowania o parametrze jest zawsze większa aniżeli „klasyczna” (choć zbiegają w )

🡪 Ale dlaczego? 🡪 I czy to „dobrze”, czy „niedobrze”?

🡪 I czy w ogóle można porównywać z ?

* + = „klasyczny”

🡪

**Testowanie pojedynczego parametru strukturalnego**

* **Bayesowskie testowanie hipotez statystycznych** = „temat rzeka”…
* **TU: rozważmy układ hipotez** – typowy na gruncie „klasycznym” ():

vs.

🡪 **Test Lindley’a**: Czy (przy zadanym poziomie p-stwa *a posteriori* ):

* + TAK: Testowana wartość parametru jest wiarygodna (wspierana przez dane), co wskazuje na zasadność
  + NIE: Testowana wartość parametru NIE jest wiarygodna (nie jest wspierana przez dane), co wskazuje na zasadność

🡪 W szczególności – **testowanie *istotności* pojedynczego parametru**:

vs.

**~~Testowanie istotności parametrów strukturalnych 1~~**

* ~~W szczególności –~~ **~~testowanie~~ *~~istotności~~* ~~pojedynczego parametru~~**~~:~~
* ~~Alternatywnie, do tego zagadnienia można podejść jako do zagadnienia~~ **~~porównywania modeli~~**~~:~~

~~vs.~~

~~⇔~~

~~vs.~~

* ~~Do zagadnienia porównywania modeli można również sprowadzić~~ **~~testowanie~~ *~~łącznej~~* ~~istotności grupy parametrów~~**~~, których wektor możemy oznaczyć jako , przy następującym podziale części deterministycznej modelu:~~

~~Wtedy~~

~~vs.~~

~~⇔~~

~~vs.~~

* **~~Wymóg~~**~~: w porównywanych modelach musi się dać obliczyć wartość 🡪 [RJn] „odpada”~~

**~~Testowanie istotności parametrów strukturalnych 2~~**

* **~~Rozważamy tu tylko przypadek [GNz]~~**
* ~~Do porównywania modeli potrzebujemy , które tu możemy oznaczyć jako i , lub – równoważnie – i , czyli wartości brzegowej gęstości obserwacji w modelach, odpowiednio: zredukowanym (, odpowiadającym ) i pełnym (, odpowiadającym )~~
* ~~Niech oznacza model regresji z danym zestawem regresorów. Ogólnie~~

~~🡪 W BMNRL z [GNz] da się ją analitycznie obliczyć, ALE mamy też inny sposób: , wobec czego~~

* ~~W przypadku~~ **~~[GNz]~~** ~~zarówno , jak i dane są dokładnymi wzorami analitycznymi, więc powyższe działanie da się wykonać:~~

~~🡪 W parametry nie są obecne, co oznacza, że ich wartości w powyższych zapisach możemy ustalić na dowolnym poziomie, a i tak się skrócą. Proponuje się przyjąć np. lub , co powinno nas uchronić przed otrzymaniem numerycznego zero dla~~

**~~Testowanie istotności parametrów strukturalnych 3~~**

* **~~W MS Excel~~** ~~– brak implementacji wielowymiarowego rozkładu normalnego, ALE:~~
  + ~~Przy typowym założeniu niezależności~~ *~~a priori~~* ~~poszczególnych parametrów strukturalnych:~~

~~potrzeba zadać diagonalną macierz , a po temu – także diagonalną macierz , gdyż~~

~~Wtedy też:~~

~~🡪 Gęstość 1-wymiarowego rozkładu normalnego w argumencie :~~

~~= ROZKŁ.NORMALNY(~~*~~x~~*~~; ; ; 0)~~

* + ~~Podobnie możemy zdekomponować rozkład próbkowy (funkcję wiarygodności), ponieważ – zgodnie z założeniami KMNRL – obserwacje są między sobą niezależne (przy ustalonym ):~~
  + ~~Dla rozkładu~~ *~~a posteriori~~*~~:~~

**~~Testowanie istotności parametrów strukturalnych 4~~**

* ~~Obliczywszy (czyli ) oraz (czyli ), możemy wyznaczyć iloraz szans~~ *~~a posteriori~~*

~~który przy jednakowych p-stwach~~ *~~a priori~~* ~~obydwu testowanych modeli redukuje się do czynnika Bayesa~~

~~informującego – w odniesieniu do hipotez – o tym, ilokrotnie bardziej/mniej wiarygodna jest hipoteza (nieistotność parametrów zgromadzonych w – lub pojedynczego ) w stosunku do (istotność). Tę kwestię możemy również sformułować w kategoriach redukcji modelu do : orzeka, że taka redukcja jest zasadna, podczas gdy opowiada się za modelem pełnym.~~

* ~~Zauważmy, że możemy wyznaczyć dla dowolnego modelu – tj. z dowolnym zestawem zmiennych objaśniających ⇒ możemy porównać moc wyjaśniającą ich wszystkich (~~**~~ranking~~**~~)~~
* ~~Równanie: ; jeden z parametrów (np. ) jest wyrazem wolnym – zawsze uwzględniany w modelu; pozostaje regresorów, z których możemy tworzyć różne zestawy~~

~~🡪 Ile jest możliwych takich zestawów:~~

**~~Testowanie istotności parametrów strukturalnych 5~~**

* ~~Skale do porównywania testowanych modeli (umowne!)~~

~~🡪 dla (czyli )~~

~~🡪 określają, z jaką siłą dane świadczą przeciw modelowi (odrzucają go) na rzecz~~

|  |  |
| --- | --- |
| ~~H. Jeffreys (1967)~~ | ~~R.E. Kass i A.E. Raftery (1995)~~ |
| * ~~0 < log~~~~10~~*~~BF~~*~~01~~ ~~≤ 0,5 🡪 słabo~~ * ~~0,5 < log~~~~10~~*~~BF~~*~~01~~ ~~≤ 1 🡪 solidnie~~ * ~~1 < log~~~~10~~*~~BF~~*~~01~~ ~~≤ 1,5 🡪 silnie~~ * ~~1,5 < log~~~~10~~*~~BF~~*~~01~~ ~~≤ 2 🡪 bardzo silnie~~ * ~~2 < log~~~~10~~*~~BF~~*~~01~~ ~~🡪 zdecydowanie~~ | * ~~0 < log~~~~10~~*~~BF~~*~~01~~ ~~≤ 1 🡪 słabo~~ * ~~1 < log~~~~10~~*~~BF~~*~~01~~ ~~≤ 3 🡪 dane pozytywnie świadczą przeciw~~ *~~M~~*~~0~~ * ~~3 < log~~~~10~~*~~BF~~*~~01~~ ~~≤ 5 🡪 silnie~~ * ~~5 < log~~~~10~~*~~BF~~*~~01~~ ~~🡪 bardzo silnie~~ |