Stopy zbieżności

- Stopę zbieżności ciągu zmiennych losowych a_n, takiego, że $a_n \xrightarrow{p} 0$ oznaczamy jako $a_n = o_p(1)$

- Jeśli $\frac{a_n}{n^{\boldsymbol{\alpha}}} \xrightarrow{p} 0$, to $a_n = o_p(n^{\boldsymbol{\alpha}})$ i mówimy a_n zbiega według prawdopodobieństwa szybciej niż $n^{\boldsymbol{\alpha}}$

- Zauważmy, że jeśli $a_n \xrightarrow{p} 0$, to $\left(\frac{n^{\boldsymbol{\alpha}} a_n}{n^{a}}\right) \xrightarrow{p} 0$ a więc $a_n = o_p(1) \Longrightarrow n^{\boldsymbol{\alpha}} a_n = o_p(n^{\boldsymbol{\alpha}})$

- Stopę zbieżności ciągu zmiennych losowych a_n, takiego, że $\mathrm{E}(a_n) < \infty$ i $\mathrm{Var}(a_n) < \infty$ dla wszytkich n oraz $\mathrm{plim}\,[\mathrm{E}(a_n)] < \infty$ i $\mathrm{plim}\,\mathrm{Var}(a_n) < \infty$, oznaczamy jako $a_n = O_p(1)$ (pojęcie te definiuje się czasami nieco ogólniej)

- Jeśli $\frac{a_n}{n^{\boldsymbol{\alpha}}} = O_p(1)$, to $a_n = O_p(n^{\boldsymbol{\alpha}})$ i mówimy a_n zbiega conajmniej tak szybko jak $n^{\boldsymbol{\alpha}}$

- Podobnie jak w poprzednim przypadku, jeśli $a_n = O_p(1) \implies n^{\boldsymbol{\alpha}} a_n = O_p(n^{\boldsymbol{\alpha}})$

- Dodatkowo $a_n = o_p(1)$ i $b_n = O_p(1)$, to $a_n b_n = o_p(1)$

Estymatory M

- Estymator M $\widehat{\boldsymbol{\theta}}$ jest rozwiązaniem problemu minimalizacji

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{N}\sum_{i=1}^{N} q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)$$

dla pewnej funkcji celu $q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)$

- Zakładamy, że wektor wektor parametrów $\boldsymbol{\theta}_0$ jest jedynym rozwiązaniem tego problemu w populacji - mówimy wtedy problem, że $\boldsymbol{\theta}$ jest zidentyfikowana

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathrm{E}\left[q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)\right]$$

- Z Prawa Wielkich Liczb wiemy, że

$$\frac{1}{N}\sum_{i=1}^{N} q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right) \xrightarrow{p} \mathrm{E}\left[q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)\right]$$

i jeśli $\widehat{\boldsymbol{\theta}}$ minimalizuje lewą stronę a $\boldsymbol{\theta}_0$ prawą stronę, to wydaje się logiczne, by $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$

Przykład Nieliniowa Metoda Najmniejszych Kwadratów

$$y_i = m\left(\boldsymbol{x}_i, \boldsymbol{\theta}_0\right) + u_i$$

$$\mathrm{E}\left(u_i | \, \boldsymbol{x}_i\right) = 0$$

Funkcją celu dla NMNK jest minimalizacja sumy kwadratów reszt $q_i\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right)$

$$q\left(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}\right) = \left[y_i - m\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right)\right]^2$$

Wartość oczekiwana q_i

$$
\begin{aligned}
\mathrm{E}\left[q\left(y_i,\boldsymbol{x}_i,\boldsymbol{\theta}\right)\mid \boldsymbol{x}_i\right] &= \mathrm{E}\left[y_i - m\left(\boldsymbol{x}_i,\boldsymbol{\theta}\right)\mid \boldsymbol{x}_i\right]^2 \\
&= \mathrm{E}\left[y_i - m\left(\boldsymbol{x}_i,\boldsymbol{\theta}_0\right) + \left[m\left(\boldsymbol{x}_i,\boldsymbol{\theta}\right) - m\left(\boldsymbol{x}_i,\boldsymbol{\theta}_0\right)\right]\mid \boldsymbol{x}_i\right]^2 \\
&= \mathrm{E}\left(u_i^2\mid \boldsymbol{x}_i\right) - 2\,\mathrm{E}\left(u_i\mid \boldsymbol{x}_i\right)\left[m\left(\boldsymbol{x}_i,\boldsymbol{\theta}\right) - m\left(\boldsymbol{x}_i,\boldsymbol{\theta}_0\right)\right] \\
&\quad + \left[m\left(\boldsymbol{x}_i,\boldsymbol{\theta}\right) - m\left(\boldsymbol{x}_i,\boldsymbol{\theta}_0\right)\right]^2 \\
&= \mathrm{E}\left(u_i^2\mid \boldsymbol{x}_i\right) + \left[m\left(\boldsymbol{x}_i,\boldsymbol{\theta}\right) - m\left(\boldsymbol{x}_i,\boldsymbol{\theta}_0\right)\right]^2
\end{aligned}
$$

a więc

$$
\mathrm{E}\left[q\left(y_i,\boldsymbol{x}_i,\boldsymbol{\theta}\right)\mid \boldsymbol{x}_i\right] > \mathrm{E}\left[q\left(y_i,\boldsymbol{x}_i,\boldsymbol{\theta}_0\right)\mid \boldsymbol{x}_i\right] = \mathrm{E}\left(u_i^2\mid \boldsymbol{x}_i\right) \text{ dla } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0
$$

Jeśli policzymy teraz wartości oczekiwane względem $\boldsymbol{x}_i$ to otrzymamy bezwarunkowe wartości oczekiwane:

$$
\mathrm{E}\left[q\left(y_i,\boldsymbol{x}_i,\boldsymbol{\theta}\right)\right] > \mathrm{E}\left[q\left(y_i,\boldsymbol{x}_i,\boldsymbol{\theta}_0\right)\right] \text{ dla } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0
$$

Istotnie więc w $\boldsymbol{\theta}_0$ jest minimum wartości oczekiwanej funkcji celu.

Jedynym założeniem, które było nam potrzebne do udowodnienia tego, że możemy zastosować estymator M było $\mathrm{E}\left(u_i | \boldsymbol{x}_i\right) = 0$. Estymator M można więc wyprowadzić przy znacznie słabszych sałożeniach niż estymator MNW dla tego modelu.

Jednostajna zbieżność i zgodność

- Aby dowieść, zgodność estymatora musimy dowieść, że $\frac{1}{N}\sum_{i=1}^{N} q(\boldsymbol{w}_i, \boldsymbol{\theta})$ jest jednostajnie zbieżne do swojej wartości oczekiwanej (co jest silniejszym typem zbieżności niż normalna zbieżność):

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{N}\sum_{i=1}^{N} q(\boldsymbol{w}_i, \boldsymbol{\theta}) - \mathrm{E}\left[q(\boldsymbol{w}_i, \boldsymbol{\theta})\right] \right| \xrightarrow{p} 0$$

- **Intuicja**: jednostajna zbieżność zachodzi, gdy maksymalna różnica między funkcjami dąży według prawdopodobieństwa do zera

Przykład Nieliniowa Metoda Najmniejszych Kwadratów

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left| \frac{1}{N} \sum_{i=1}^{N} \left[m\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right) + u_i \right] - \mathrm{E}\left[m\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right) + u_i \right] \right| \xrightarrow{p} 0$$

będzie spełnione jeśli $\frac{1}{N}\sum_{i=1}^{N} u_i \xrightarrow{p} 0$ i $\frac{1}{N}\sum_{i=1}^{N} m\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right) \xrightarrow{p} \mathrm{E}\left[m\left(\boldsymbol{x}_i, \boldsymbol{\theta}\right)\right]$, dla każdego $\boldsymbol{\theta}$.

- Jeśli zachodzi jednostajna zbieżność, to

$$
\begin{aligned}
\operatorname{plim} \widehat{\boldsymbol{\theta}} &= \operatorname{plim} \left[\arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right) \right] \\
&= \arg\min_{\boldsymbol{\theta}} \left[\operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right) \right] \\
&= \arg\min_{\boldsymbol{\theta}} \left\{ \mathrm{E}\left[q\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right)\right]\right\} = \boldsymbol{\theta}_0
\end{aligned}
$$

przy czym jednostajna zbieżności umożliwia policzenie granicy argumentu ($\operatorname{plim}\left[\arg\min_{\boldsymbol{\theta}}(\cdot)\right]$) jako argumentu granicy ($\arg\min_{\boldsymbol{\theta}}\left[\operatorname{plim}(\cdot)\right]$)

- Estymator M jest zgodny.

Dowód. Z definicji jednostajnej zbieżności oraz zbieżności funkcji celu mamy:

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{i=1}^{N}q\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)-\mathrm{E}\left[q\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right]\right|+\left|\mathrm{E}\left[q\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)\right]-\frac{1}{N}\sum_{i=1}^{N}q\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)\right|\xrightarrow{p}0$$

Korzystając z tego, że $|x_1|+|x_2|\geq|x_1+x_2|\geq 0$ mamy

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{i=1}^{N}q\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)-\frac{1}{N}\sum_{i=1}^{N}q\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)+\mathrm{E}\left[q\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)\right]-\mathrm{E}\left[q\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right]\right|\xrightarrow{p}0$$

ale z założenia, że $\mathrm{E}\left[q\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)\right]$ jest minimum wynika, że $\mathrm{E}\left[q\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right]-\mathrm{E}\left[q\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)\right]=\boldsymbol{\alpha}\geq 0$ a więc

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{i=1}^{N}q\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)-\frac{1}{N}\sum_{i=1}^{N}q\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)-\boldsymbol{\alpha}\right|\xrightarrow{p}0$$

Z tego z kolei wnioskujemy, że

$$\frac{1}{N}\sum_{i=1}^{N} q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right) - \frac{1}{N}\sum_{i=1}^{N} q\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right) \xrightarrow{p} \boldsymbol{\alpha} \geq 0 \text{ dla każdego } \boldsymbol{\theta}$$

przy czym równość zachodzi jedynie dla $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Wnioskujemy z tego, że w granicy $\frac{1}{N}\sum_{i=1}^{N} q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)$ osiąga minimum dla $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. ■

- Jeśli $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ i funkcja $r\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)$ jest jednostajnie zbieżna do $\mathrm{E}\left[r\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)\right]$, to

  $$\frac{1}{N}\sum_{i=1}^{N} r\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right) \xrightarrow{p} \mathrm{E}\left[r\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right)\right]$$

  i w związku z tym $\frac{1}{N}\sum_{i=1}^{N} r\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right)$ jest zgodnym estymatorem $\mathrm{E}\left[r\left(w_i, \widehat{\boldsymbol{\theta}}\right)\right]$

  Dowód. Z definicji jednostajnej zbieżności oraz zbieżności funkcji celu mamy:

$$\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{N}\sum_{i=1}^{N} r\left(\boldsymbol{w}_i,\widehat{\boldsymbol{\theta}}\right)-\mathrm{E}\left[r\left(\boldsymbol{w}_i,\widehat{\boldsymbol{\theta}}\right)\right]\right| \xrightarrow{p} 0$$

Z twierdzenia Cramera wynika

$$\mathrm{E}\left[r\left(\boldsymbol{w}_i,\widehat{\boldsymbol{\theta}}\right)\right] \xrightarrow{p} \mathrm{E}\left[r\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right]$$

Z kolei z jednostajnej zbieżności wnioskujemy, że

$$\frac{1}{N}\sum_{i=1}^{N} r\left(\boldsymbol{w}_i,\widehat{\boldsymbol{\theta}}\right) \xrightarrow{p} \mathrm{E}\left[r\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right]$$

■

Asymptotyczna normalność estymatorów M

- Założenia:

  1. spełnione są warunki, dla których estymator M jest zgodny $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$
  2. poszczególne $\boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})$ są niezależne i mają identyczne rozkłady bezwarunkowe
  3. gradient $\frac{1}{N}\sum_{i-1}^{N} \boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})\, \boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})'$ i hessjan $\frac{1}{N}\sum_{i-1}^{N} \boldsymbol{H}(\boldsymbol{w}_i, \boldsymbol{\theta})$ są jednostajnie zbieżne odpowiednio do $\mathrm{E}\left[\boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})\, \boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})'\right]$ i $\mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}_i, \boldsymbol{\theta})\right]$

  gdzie $\boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})$ jest pierwszą pochodną $q(\boldsymbol{w}_i, \boldsymbol{\theta})$

- Estymator $\widehat{\boldsymbol{\theta}}$ minimalizuje wartość oczekiwaną $q(\boldsymbol{w}, \boldsymbol{\theta})$

- Jeśli $q(\boldsymbol{w}, \boldsymbol{\theta})$ jest ciągle różniczkowalna, to pierwsza pochodna $q(\boldsymbol{w}, \boldsymbol{\theta})$ powinna być równa 0 dla $\widehat{\boldsymbol{\theta}}$

$$\sum_{i-1}^{N} \boldsymbol{s}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right) = 0$$

- Funkcja $\boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta})$ jest gradientem elementu funkcji $q(\boldsymbol{w}_i, \boldsymbol{\theta})$ celu (*score*)

- Rozwinięcie Taylora $\sum_{i-1}^{N} \boldsymbol{s}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right)$ wokół $\boldsymbol{\theta}_0$ daje

$$\sum_{i-1}^{N} \boldsymbol{s}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right) = \sum_{i-1}^{N} \boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta}_0) + \left(\sum_{i-1}^{N} \boldsymbol{H}_i\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + o_p(1)$$

gdzie $\boldsymbol{H}_i = \boldsymbol{H}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right)$ jest hessjanem funkcji celu. Zbieżność reszty wynika z tego, że $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$.

- Jak już wspomniano $\sum_{i-1}^{N} \boldsymbol{s}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right) = 0$

$$\sum_{i-1}^{N} \boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta}_0) = -\left(\sum_{i-1}^{N} \boldsymbol{H}_i\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + o_p(1)$$

- Z Prawa Wielkich Liczb wynika, że $N^{-1}\sum_{i-1}^{N} \boldsymbol{H}_i \xrightarrow{p} \mathrm{E}[\boldsymbol{H}(\boldsymbol{w},\boldsymbol{\theta}_0)] = \boldsymbol{A}_0 = O_p(1)$ (zakładamy dodatkowo, że $\boldsymbol{A}_0$ jest nieosobliwa).

- Mnożąc obie strony przez $\frac{1}{N}$ uzyskujemy

$$\frac{1}{N}\sum_{i-1}^{N} \boldsymbol{s}(\boldsymbol{w}_i, \boldsymbol{\theta}_0) = -\left(\frac{1}{N}\sum_{i-1}^{N} \boldsymbol{H}_i\right)\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) + o_p\left(\frac{1}{N}\right)$$

dzieląc obie strony przez $\frac{1}{N}\sum_{i-1}^{N}\boldsymbol{H}_i = O_p(1)$, przenosząc na drugą stronę i mnożąc przez $\sqrt{N}$ uzyskujemy

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)=-\left(\frac{1}{N}\sum_{i-1}^{N}\boldsymbol{H}_i\right)^{-1}\left[-\frac{1}{\sqrt{N}}\sum_{i-1}^{N}\boldsymbol{s}\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)\right]+o_p\left(\frac{1}{\sqrt{N}}\right)$$

- A więc

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)=\boldsymbol{A}_0^{-1}\left[-\frac{1}{\sqrt{N}}\sum_{i-1}^{N}\boldsymbol{s}\left(\boldsymbol{w}_i,\boldsymbol{\theta}_0\right)\right]+o_p\left(\frac{1}{\sqrt{N}}\right)$$

- Zauważmy, że przy spełnionych warunkach regularności:

$$\frac{\partial}{\partial\boldsymbol{\theta}}\mathrm{E}\left[q\left(\boldsymbol{w},\boldsymbol{\theta}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=\mathrm{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}}q\left(\boldsymbol{w},\boldsymbol{\theta}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}=\mathrm{E}\left[\boldsymbol{s}\left(\boldsymbol{w},\boldsymbol{\theta}_0\right)\right]=0$$

- Ponieważ zgodnie z założeniami $q\left(\boldsymbol{w}_i, \boldsymbol{\theta}\right)$ są niezależne i mają identyczne rozkłady więc $\boldsymbol{s}\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right)$ spełnia Centralnego Twierdzenia Granicznego i

$$\frac{1}{\sqrt{N}} \sum_{i-1}^{N} \boldsymbol{s}\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right) \xrightarrow{D} N\left(0, \boldsymbol{B}_0\right)$$

gdzie

$$\boldsymbol{B}_0 = \mathrm{E}\left[\boldsymbol{s}\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right) \boldsymbol{s}\left(\boldsymbol{w}_i, \boldsymbol{\theta}_0\right)'\right]$$

- W rezultacie

$$\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{D} N\left(0, \boldsymbol{A}_0^{-1} \boldsymbol{B}_0 \boldsymbol{A}_0^{-1}\right)$$

- Rozkład estymatora M jest normalny!

- Aasymptotyczna aproksymacja wariancji estymatora może być znaleziona jako

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \overset{a}{\sim} N\left(\mathbf{0}, \frac{1}{N}\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\right)$$

Estymatory macierzy wariancji dla estymatorów M

- Często trudno jest bezpośrednio znaleźć $\boldsymbol{B}_0 = \mathrm{E}\left[\boldsymbol{s}(\boldsymbol{w}, \boldsymbol{\theta}_0)\, \boldsymbol{s}(\boldsymbol{w}, \boldsymbol{\theta}_0)'\right]$ i $\boldsymbol{A}_0 = \mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}, \boldsymbol{\theta}_0)\right]$.

- Jednak używając poprzednio wprowadzonego Lematu:

$$\widehat{\boldsymbol{A}} = N^{-1} \sum_{i-1}^{N} \boldsymbol{H}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right) = N^{-1} \sum_{i-1}^{N} \widehat{\boldsymbol{H}}_i \xrightarrow{p} \mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}, \boldsymbol{\theta}_0)\right] = \boldsymbol{A}_0$$

- Wady tego estymatora:
  - konieczne policzenie macierzy drugich pochodnych
  - estymator nie zawsze musi być dodatnio określony

- W wielu zastosowanich ekonomicznych wnioskowanie prowadzimy warunkowo wględem $\boldsymbol{x}_i$ i liczymy w związku z tym warunkową macierz:

$$\boldsymbol{A}_i = \boldsymbol{A}\left(\boldsymbol{x}_i,\boldsymbol{\theta}_0\right) = \mathrm{E}\left[\left.\boldsymbol{H}\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right|\boldsymbol{x}_i\right]$$

- Zgodnie z prawem iterowanych wartosci oczekiwanych $\mathrm{E}\left[\boldsymbol{A}\left(\boldsymbol{x}_i,\boldsymbol{\theta}_0\right)\right] = \mathrm{E}\left[\boldsymbol{H}\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right] = \boldsymbol{A}_0$

- Dalej

$$\widehat{\boldsymbol{A}} = N^{-1}\sum_{i-1}^{N}\boldsymbol{A}\left(\boldsymbol{x}_i,\widehat{\boldsymbol{\theta}}\right) = N^{-1}\sum_{i-1}^{N}\widehat{\boldsymbol{A}}_i \xrightarrow{p} \boldsymbol{A}_0$$

- Ten estymator ma sens jeśli można łatwo uzyskać $\mathrm{E}\left[\left.\boldsymbol{H}\left(\boldsymbol{w}_i,\boldsymbol{\theta}\right)\right|\boldsymbol{x}_i\right]$. W wielu przypadkach można pokazać, że estymator ten jest dodatnio określony.

- Estymator B_0można uzyskać w sposób analogiczny:

$$\widehat{\boldsymbol{B}} = N^{-1} \sum_{i-1}^{N} \boldsymbol{s}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right) \boldsymbol{s}\left(\boldsymbol{w}_i, \widehat{\boldsymbol{\theta}}\right)' = N^{-1} \sum_{i-1}^{N} \widehat{\boldsymbol{s}}_i \widehat{\boldsymbol{s}}_i' \xrightarrow{p} \boldsymbol{B}_0$$

- W rezultacie zgodnym estymatorem wariancji $\sqrt{N}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$ będzie

$$\widehat{\boldsymbol{V}} = \widehat{\boldsymbol{A}}^{-1} \widehat{\boldsymbol{B}} \widehat{\boldsymbol{A}}^{-1}$$

- A asymptotyczną aproksymacją wariancji $\widehat{\boldsymbol{\theta}}$ jest

$$Avar\left(\widehat{\boldsymbol{\theta}}\right) = \widehat{\boldsymbol{A}}^{-1} \widehat{\boldsymbol{B}} \widehat{\boldsymbol{A}}^{-1} \Big/ N$$

- W wielu przypadkach można uzyskać znaczne uproszczenie, jeśli

$$\mathrm{E}\left[\boldsymbol{s}\left(\boldsymbol{w},\boldsymbol{\theta}\right)\boldsymbol{s}\left(\boldsymbol{w},\boldsymbol{\theta}\right)'\right] = -\boldsymbol{\sigma}_0^2\,\mathrm{E}\left[\boldsymbol{H}\left(\boldsymbol{w},\boldsymbol{\theta}_0\right)\right] \qquad (*)$$

albo $\boldsymbol{B}_0 = \boldsymbol{\sigma}^2\boldsymbol{A}_0$

- W tym szczególnym (choć częstym - tak jest np. w MNW dla $\boldsymbol{\sigma}_0^2 = 1$) przypadku:

$$\widehat{\boldsymbol{V}} = -\,\widehat{\boldsymbol{\sigma}}^2\widehat{\boldsymbol{A}}^{-1} = \widehat{\boldsymbol{B}}^{-1}$$

- Macierz warinacji kowariancji $Av\widehat{a}r\left(\widehat{\boldsymbol{\theta}}\right) = \widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{A}}^{-1}\big/N$ jest niekiedy nazywana odporną macierz wariancji kowariancji (*robust*), ponieważ daje prawidłowe oszacowania nawet wtedy, gdy nie jest spełnione założenie *

Przykład Szacowanie macierzy wariancji kowariancji w MNK w przypadku występowania heteroskedastyczności.

Standardowa postać macierzy wariancji kowariancji estymator MNK została wyprowadzona przy założeniu braku heteroskedastyczności ($KMRL$). Dowiedliśmy wcześniej, że estymator M oparty na minimalizacji sumy kwadratów reszt dla modelu postaci $y_i = m(\boldsymbol{x}_i, \boldsymbol{\theta}_0) + u_i$ jest zgodny jeśli tylko $\mathrm{E}(u_i|\, \boldsymbol{x}_i) = 0$. Odnosi się to oczywiście także do modelu do modelu liniowego $m(\boldsymbol{x}_i, \boldsymbol{\beta}) = \boldsymbol{x}_i\boldsymbol{\beta}$, $q(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}) = (y_i - \boldsymbol{x}_i\boldsymbol{\beta})^2$. Zakładmy dodatkowo, że $\mathrm{Var}(u_i) = \boldsymbol{\sigma}^2$, że u_i, u_j są niezależne i u_i ma ten sam rozkład dla każdego i.

Różnica w stosunku do KMRL. Nie zakładamy braku zależności między wariancją a zmiennymi objaśniającymi. $\mathrm{Var}(u_i|\, \boldsymbol{x}_i)$ może nie być stała dla każdego i. Zakładamy jedynie, że $\mathrm{E}_x[\mathrm{Var}(u_i|\, \boldsymbol{x}_i)] = \mathrm{Var}(u_i) = \boldsymbol{\sigma}^2$.

Dla modelu liniowego wektor score i oszacowania z parametru $\boldsymbol{\beta}$

oznaczonego $\boldsymbol{b}$ mamy

$$\begin{aligned} \boldsymbol{s}\left(y_i, \boldsymbol{x}_i, \boldsymbol{b}\right) &= 2\boldsymbol{x}_i'\boldsymbol{x}_i\,\boldsymbol{b} - 2\boldsymbol{x}_i' y_i = \\ &= -\boldsymbol{x}_i'\left(y_i - \boldsymbol{x}_i\boldsymbol{b}\right) = -2\boldsymbol{x}_i' e_i \end{aligned}$$

W związku z tym oszacowaniem $\boldsymbol{B}$ jest równe $\widehat{\boldsymbol{B}} = 4N^{-1}\sum_{i-1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i e_i^2$. Hessian funkcji celu jest równy

$$\boldsymbol{H}\left(y_i, \boldsymbol{x}_i, \boldsymbol{\beta}\right) = 2\boldsymbol{x}_i'\boldsymbol{x}_i$$

a więc oszacowaniem macierzy $\boldsymbol{A}_0$ będzie

$$\widehat{\boldsymbol{A}} = N^{-1}\sum_{i-1}^{N} 2\boldsymbol{x}_i'\boldsymbol{x}_i = 2\boldsymbol{X}'\boldsymbol{X}$$

W rezultacie oszacowaniem macierzy wariancji kowariancji będzie

$$Avar\left(\boldsymbol{b}\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\sum_{i-1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i e_i^2\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

a więc macierz odporna White'a.

Testowanie hipotez

- Podobnie jak w przypadku MNW dla estymatorów M można zastosować odpowiedniki testów Walda, mnożnikow Lagrange'a LM i ilorazu wiarogodności LR.

- Przy ogólnej hipotezie nieliniowej $\boldsymbol{h}(\boldsymbol{\theta}) = 0$ i założeniu, że $\boldsymbol{H}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ ma pełen rząd

- Test Walda ma postać

$$W = \widehat{\boldsymbol{h}}' \left[\widehat{\boldsymbol{H}} \widehat{\boldsymbol{V}} \widehat{\boldsymbol{H}}' \right]^{-1} \widehat{\boldsymbol{h}} \xrightarrow{D} \boldsymbol{\chi}_g^2$$

gdzie $\widehat{\boldsymbol{V}}$ jest asympotyczną wariancją $\widetilde{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{h}} = \boldsymbol{h}\left(\widehat{\boldsymbol{\theta}}\right)$, $\widehat{\boldsymbol{H}} = \boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}\right)$

- Test mnożników Lagrange'a

$$\left(\sum_{i=1}^{N}\widehat{\boldsymbol{s}}_i\right)'\widehat{\boldsymbol{A}}^{-1}\widehat{\boldsymbol{H}}'\left(\widehat{\boldsymbol{H}}\widehat{\boldsymbol{V}}\widehat{\boldsymbol{H}}'\right)^{-1}\widehat{\boldsymbol{H}}\widehat{\boldsymbol{A}}^{-1}\left(\sum_{i=1}^{N}\widehat{\boldsymbol{s}}_i\right)\Big/ N \xrightarrow{d} \boldsymbol{\chi}_Q^2$$

gdzie wszystkie $\widehat{\boldsymbol{s}}_i = \boldsymbol{s}_i\left(\widehat{\boldsymbol{\theta}}_R\right)$, $\widehat{\boldsymbol{H}} = \boldsymbol{H}\left(\widehat{\boldsymbol{\theta}}_R\right)$.

- Test QLR (quasi iloraz wiarygodności) ma postać

$$QLR = 2\left[\sum_{i-1}^{N} q\left(\boldsymbol{w}_i,\widehat{\boldsymbol{\theta}}_R\right) - \sum_{i-1}^{N} q\left(\boldsymbol{w}_i,\widehat{\boldsymbol{\theta}}\right)\right] \xrightarrow{d} \boldsymbol{\chi}_Q^2$$

gdzie $\widehat{\boldsymbol{\theta}}_R$ jest estymatorem z ograniczeniami $\widehat{\boldsymbol{\theta}}$ bez ograniczeń

Estymatory MNW i pseudo MNW

- W przypadku modelu estymowanego MNW funkcja gęstości musimy znać warunkową funkcję gęstości dla każdej obserwacji

$$f\left(y_i|\,\boldsymbol{x}_i;\boldsymbol{\theta}\right)=p_0\left(y_i|\,\boldsymbol{x}_i\right)$$

- Estymatory MNW są specjalnym przypadkiem estymatorów M, ponieważ

$$\mathrm{E}\left[\ell_i\left(\boldsymbol{\theta}_0\right)|\,\boldsymbol{x}_i\right]\geq\mathrm{E}\left[\ell_i\left(\boldsymbol{\theta}\right)|\,\boldsymbol{x}_i\right]\text{ dla każdego }\boldsymbol{\theta}\in\boldsymbol{\Theta}$$

gdzie $\ell_i\left(\boldsymbol{\theta}\right)=\ell_i\left(\boldsymbol{y}_i,\boldsymbol{x}_i,\boldsymbol{\theta}\right)=\log\left[f\left(y_i|\,\boldsymbol{x}_i;\boldsymbol{\theta}\right)\right]$

Dowód. Z własności prawdopodobieństwa $\int_Y f\left(y_i|\, \boldsymbol{x}_i; \boldsymbol{\theta}_0\right) dy_i = 1$. Z warunkowej wersji nierówności Jensena

$$
\begin{aligned}
\mathrm{E}\left[\log\left(\frac{f\left(y_i|\, \boldsymbol{x}_i; \boldsymbol{\theta}\right)}{f\left(y_i|\, \boldsymbol{x}_i; \boldsymbol{\theta}_0\right)}\right)\middle|\, \boldsymbol{x}_i\right] &\leq \log\left\{\mathrm{E}\left[\frac{f\left(y_i|\, \boldsymbol{x}_i, \boldsymbol{\theta}\right)}{f\left(y_i|\, \boldsymbol{x}_i, \boldsymbol{\theta}_0\right)}\middle|\, \boldsymbol{x}_i\right]\right\} = \\
&= \log\left\{\int_Y \frac{f\left(y_i|\, \boldsymbol{x}_i, \boldsymbol{\theta}\right)}{f\left(y_i|\, \boldsymbol{x}_i, \boldsymbol{\theta}_0\right)} f\left(y_i|\, \boldsymbol{x}_i, \boldsymbol{\theta}_0\right) dy_i\right\} \\
&= \log\left[\int_Y f\left(y_i|\, \boldsymbol{x}_i; \boldsymbol{\theta}\right) dy_i\right] = 0
\end{aligned}
$$

i z definicji $\ell_i\left(\boldsymbol{\theta}\right)$

$$
\mathrm{E}\left[\ell_i\left(\boldsymbol{\theta}_0\right)|\, \boldsymbol{x}_i\right] \geq \mathrm{E}\left[\ell_i\left(\boldsymbol{\theta}\right)|\, \boldsymbol{x}_i\right]
$$

a więc w $\boldsymbol{\theta}_0$ wartość oczekiwana funkcji celu przyjmuje wartość *maksymalną*
■

- W tym przypadku *maksymalizujemy* funkcję celu

$$\max_{\boldsymbol{\theta}\in\Theta} \sum_{i=1}^{N} \ell_i \left(\boldsymbol{\theta}\right)$$

- Założenie, przy którym wyprowadza się standardowe własności estymatora MNW mówi, że

$$L\left(\boldsymbol{\theta}\right) = f\left(\boldsymbol{y} \middle| \boldsymbol{x};\boldsymbol{\theta}\right) = \prod_{i=1}^{N} f\left(y_i \middle| \boldsymbol{x}_i, \boldsymbol{\theta}_0\right)$$

  jest poprawną funkcją wiarygodności dla całej próby.

- Implikuje to, że poszczególne obserwacje są niezależne co często nie jest prawdą.

- Zauważmy, że do zbieżności estymatora M (a więc i estymatora MNW) wystarczy, by poprawnie była wyspecyfikowana funkcja gęstości dla *pojedynczego* zdarzenia

- **Uwaga**: w dowodzie dla $\mathrm{Var}\left(\frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) = -\mathrm{E}\left(\frac{\partial^2\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) = \boldsymbol{I}(\boldsymbol{\theta})$ używaliśmy założenia, że funkcja wiarygodności jest poprawnie wyspecyfikowana, tylko dla tego przypadku macierzą warinacji kowariancji będzie $\boldsymbol{I}(\boldsymbol{\theta})^{-1}$.

- **Wniosek**:
  - jeśli poprawnie wyspecyfikowana jest jedynie funkcja gęstości dla pojedynczego zdarzenia to estymator MNW jest dalej zgodny
  - aby poprawnie wyestymować macierz wariancji estymatora należy zastosować wersję odporną macierzy wariancji estymatora M

Przykład (Próbkowanie warstwami) Załóżmy, że chcemy wyestymować model probitowy dla próby, która została wylosowana w sposób następujący:

najpierw losowano gospodarstwa a później zadawano pytania wszystkim członkom gospodarstwa (tak losowana jest np. próba BAEL). Załóżmy dla uproszczenia, że badano jedynie gospodarstwa k osobowe. Problem estymacji w tym przypadku związany jest z ewentualnym zależnością między elementami losowymi wewnątrz gospodarstwa - poszczególne obserwacje mogą być zależne. Załóżmy, że model ma następującą postać:

$$\begin{aligned} y_{ij}^* &= \boldsymbol{x}_{ij}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{ij} \\ y_{ij} &= \begin{cases} 1 & \text{dla} \quad y_{ij}^* > 0 \\ 0 & \text{dla} \quad y_{ij}^* \leq 0 \end{cases} \\ \boldsymbol{\varepsilon}_i &\sim N(\mathbf{0}, \boldsymbol{\Sigma}) \end{aligned}$$

gdzie i jest indeksem gospodarstwa, j jest indeksem osoby wewnątrz, ε_{ij} jest czynnikiem losowym. Zakładamy, że czynniki losowe mogą być skorelowane wewnątrz gospodarstwa $\mathrm{E}(\varepsilon_i \varepsilon_i') = \boldsymbol{\Sigma} \neq \boldsymbol{\sigma}^2 I$ ale są niezależne dla różnych gospodarstw $\mathrm{E}\left(\varepsilon_i \varepsilon_j'\right) = 0$. Sformułowanie pełnej funkcji wiarygodności dla tego problemu byłoby bardzo trudne, ponieważ trudne jest policzenie dystrybuanty wielowymiarowego rozkladu normalnego. Można jednak zauważyć, że dla każdej z k obserwacji spełnione jest, że

$$\mathrm{E}\left[\ell_{ij}\left(\boldsymbol{\theta}_0\right)\mid y_i, \boldsymbol{x}_i\right] \geq \mathrm{E}\left[\ell_{ij}\left(\boldsymbol{\theta}\right)\mid y_i, \boldsymbol{x}_i\right]$$

Dalej suma $\ell_i\left(\boldsymbol{\theta}_0\right) = \sum_{j=1}^{k} \ell_{ij}\left(\boldsymbol{\theta}_0\right)$ będzie też maksymalizowane przez $\boldsymbol{\theta}_0$.

$$\mathrm{E}\left[\ell_{i}\left(\boldsymbol{\theta}_0\right)\mid y_i, \boldsymbol{x}_i\right] \geq \mathrm{E}\left[\ell_{i}\left(\boldsymbol{\theta}_0\right)\mid y_i, \boldsymbol{x}_i\right]$$

Z kolei z racji na założony brak korelacji między obserwacjami dla różnych gospodarstw, poszczególne $\ell_i\left(\boldsymbol{\theta}\right)$ będą niezależne. Wynika z tego, że zgodnym estymator M oparty na funkcji celu $\ell\left(\boldsymbol{\theta}\right) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k} \ell_{ij}\left(\boldsymbol{\theta}\right)$, gdzie N jest liczbą przebadanych gospodarstw. Estymator ten będzie miał asymptotyczny rozkład normalny o wariancji:

$$\begin{aligned}\boldsymbol{A}_0 &= \mathrm{E}\left[\boldsymbol{H}\left(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}_0\right)\right] \\ \boldsymbol{B}_0 &= \mathrm{E}\left[\boldsymbol{s}\left(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}\right) \boldsymbol{s}\left(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}\right)'\right]\end{aligned}$$

gdzie $\boldsymbol{s}\left(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}\right) = \frac{\partial \sum_{j=1}^{k} \ell_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ a $\boldsymbol{H}\left(y_i, \boldsymbol{x}_i, \boldsymbol{\theta}\right) = \frac{\partial^2 \sum_{j=1}^{k} \ell_{ij}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}}$.