Теорія ймовір. та матем. статист.
Вип. 85, 2011, стор. 150–160

Teor. Ĭmovir. ta Matem. Statyst.
No. 85, 2011, pp. 150–160

ЛОГІСТИЧНА РЕГРЕСІЯ З ГОМОСКЕДАСТИЧНОЮ ПОХИБКОЮ — МОДЕЛЬ БЕРКСОНА

УДК 519.21

С. В. ШКЛЯР

Анотація. Розглянуто логістичну регресію з одним регресором, на який накладається нормально розподілена гомоскедастична похибка спостережень — модель Берксона. Дисперсія цих похибок вважається відомою. Знайдені достатні умови того, що у структурній моделі граничне рівняння для оцінки максимальної вірогідності має рівно один розв'язок, та достатні умови сильної консистентності оцінки.

Аннотация. Рассмотрена логистическая регрессия с одним регрессором, на который наложена нормально распределенная гомоскедастическая ошибка — модель Берксона. Дисперсия этих ошибок считается известной. Найдены достаточные условия того, что в структурной модели предельное уравнение для оценки максимального правдоподобия имеет ровно одно решение, и достаточные условия сильной состоятельности оценки.

Abstract. We consider Berkson model of logistic regression with one regressor and normally distributed errors in the regressor. The variance of the errors is assumed known. Sufficient conditions for uniqueness of the solution to limit estimating equation, and sufficient conditions for strong consistency of maximum likelihood estimator are found.

1. Вступ

Статистична модель. Розглянемо модель Берксона логістичної регресії з однією незалежною змінною. Нехай X^{obs} — спостережуване значення регресора. Істинне значення регресора X^{tr} за умови X^{obs} має нормальний розподіл:

$$X^{\mathrm{tr}}|X^{\mathrm{obs}} \sim N(X^{\mathrm{obs}}, \tau^2). \tag{1}$$

Відгук Y набуває значень 0 та 1, причому

$$\mathsf{P}[Y{=}1 \mid X^{\mathrm{obs}}, X^{\mathrm{tr}}] = \frac{e^{\beta_0+\beta_1 X^{\mathrm{tr}}}}{1+e^{\beta_0+\beta_1 X^{\mathrm{tr}}}}. \tag{2}$$

Припускаємо, що дисперсія τ^2 однакова для різних спостережень, тобто похибка гомоскедастична; параметр τ^2 вважаємо відомим.

Ми розглядаємо *структурну* модель, у якій спостережувані значення регресора $X_1^{\mathrm{obs}}, X_2^{\mathrm{obs}}, \ldots, X_N^{\mathrm{obs}}$ випадкові та однаково розподілені. Реалізації моделі (1)–(2) — трійки $(X_n^{\mathrm{obs}}, X_n^{\mathrm{tr}}, Y_n)$, $n \ge 1$ — незалежні та однаково розподілені.

Спостерігаються пари $(X_n^{\mathrm{obs}}, Y_n)$, $n{=}1, 2, \ldots, N$. Будемо оцінювати векторний параметр регресії $\beta = (\beta_0, \beta_1)^\top$.

2000 *Mathematics Subject Classification.* Primary 62J12; Secondary 62G20.

Ключові слова і фрази. Логістична регресія, бінарна регресія, моделі з похибками в змінних, модель Берксона, модель калібрування регресії.

Робота базується на доповіді, що була представлена на міжнародній конференції "Modern Stochastics: Theory and Applications II", що проходила 7–11 вересня 2010 г. у Київському національному университеті імені Тараса Шевченка и була присвячена ювілеям трьох видатних українських вчених: Анатолія Скорохода, Володимира Королюка та Ігоря Коваленка.

Огляд літератури. Модель Берксона для бінарної регресії розглянута в статті [2]. Але там теоретичні результати наведено не для логістичної регресії, а для пробіт-регресії.

Модель з класичною похибкою вимірювання, а не модель Берксона, розглянуто в роботах [4, 7]. У статті [4] наведено різні методи оцінювання; у статті [7] показано, що в моделі з одним регресором з класичною похибкою та нормальним розподілом регресора, якщо коефіцієнт перед X відмінний від 0, то модель ідентифіковна.

Отриманий результат. Ми знаходимо умови, за яких оцінка максимальної вірогідності (ОМВ) сильно консистентна. Доведено, що якщо оцінку означати як точку максимуму функції вірогідності на компактній множині, яка містить істинне значення параметра як внутрішню точку, то для сильної консистентності оцінки достатньо, щоб розподіл X^{obs} задовольняв наступні умови: розподіл X^{obs} не зосереджений в одній точці та $\mathsf{E}\,|X^{\mathrm{obs}}| < \infty$.

Згадаємо деякі проміжні результати. За умов, наведених у попередньому абзаці, модель ідентифіковна (це одна з умов консистентності оцінки максимальної вірогідності). Також доведено, що за цих самих умов граничне рівняння $\mathsf{E}_\beta\, S(b) = 0$ (тут $S(b)$ — градієнт логарифму функції вірогідності) має рівно один розв'язок на $\mathbb{R}^2$, а саме $b = \beta$.

Структура статті. У розділі 2 встановлено різні співвідношення між згорткою щільностей логістичного та нормального розподілів та первісною і похідними цієї згортки. У розділі 3 наведено функцію вірогідності та відповідне оціночне рівняння. У розділі 4 досліджено граничне рівняння $\mathsf{E}_\beta\, S(b) = 0$. У розділі 5 доведено основний результат — сильну консистентність оцінки максимальної вірогідності.

2. Згортка логістичного та нормального розподілів

2.1. **Означення та похідні функції $L_0(x, \sigma^2)$.** Логістичний розподіл — це розподіл, заданий функцією розподілу

$$\mathsf{P}(\lambda < x) = \frac{e^x}{e^x + 1}.$$

Щільність логістичного розподілу дорівнює $\frac{e^x}{(e^x+1)^2}$.

Розглянемо суму незалежних випадкових величин $\lambda + \xi$, де λ має логістичний розподіл, а ξ має нормальний розподіл, $\xi \sim N(0, \sigma^2)$, $\sigma^2 \ge 0$ (допускається випадок $\sigma^2 = 0$). Функцію розподілу $\lambda + \xi$ позначимо через $L_0(x, \sigma^2)$. Маємо

$$\begin{aligned} L_0\left(x, \sigma^2\right) &= \mathsf{P}(\lambda + \xi < x) = \mathsf{P}(\lambda < x - \xi) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}\, \mathsf{P}(\lambda < x - t)\, dt \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2\sigma^2}}\, \frac{e^{x-t}}{e^{x-t}+1}\, dt = \mathsf{E}\, \frac{e^{x-\xi}}{e^{x-\xi}+1}. \end{aligned}$$

Позначимо щільність розподілу $\lambda + \xi$ через $L_1(x, \sigma^2)$; взагалі, позначимо

$$L_k(x, \sigma^2) = \frac{\partial^k}{\partial x^k} L_0(x, \sigma^2), \qquad x{\in}\mathbb{R},\ k{\ge}1.$$

Диференціюючи під знаком математичного сподівання, отримаємо рівність

$$L_k(x, \sigma^2) = \mathsf{E}\, \frac{\partial^k}{\partial x^k} \left(\frac{e^{x-\xi}}{e^{x-\xi}+1} \right), \tag{3}$$

де $\xi \sim N(0, \sigma^2)$.

Зауважимо, що для всіх $x \in \mathbb{R}$

$$0 < L_0(x, \sigma^2) < 1, \qquad L_1(x, \sigma^2) > 0.$$

При фіксованому σ^2 функція $L_0(x,\sigma^2)$ строго зростає за аргументом x. Розподіл $\xi+\lambda$ симетричний, тому

$$L_0(-x,\sigma^2)+L_0(x,\sigma^2)=1.$$

З обмеженості похідних функції $\frac{e^x}{1+e^x}$ випливає обмеженість функцій $L_k(x,\sigma^2)$ при фіксованому k, зокрема

$$0<L_1(x,\sigma^2)\le\frac{1}{4},\qquad |L_2(x,\sigma^2)|\le\frac{1}{6\sqrt{3}}$$

(нерівності строгі, якщо $\sigma^2>0$).

Знайдемо похідну $\frac{\partial}{\partial t}L_0(x,t)$, де $t\ge 0$. Для цього розглянемо стандартний вінерівський процес W_t. За формулою Іто

$$d\left(\frac{e^{x+W_t}}{e^{x+W_t}+1}\right)=\frac{e^{x+W_t}}{(e^{x+W_t}+1)^2}\,dW_t+\frac{1}{2}\cdot\frac{e^{x+W_t}-e^{2x+2W_t}}{(e^{x+W_t}+1)^3}\,dt.$$

Візьмемо математичне сподівання:

$$\mathsf{E}\,d\left(\frac{e^{x+W_t}}{e^{x+W_t}+1}\right)=\frac{1}{2}\,\mathsf{E}\,\frac{e^{x+W_t}-e^{2x+2W_t}}{(e^{x+W_t}+1)^3}\,dt,$$

тобто

$$\mathsf{E}\,\frac{e^{x+W_{t_2}}}{e^{x+W_{t_2}}+1}-\mathsf{E}\,\frac{e^{x+W_{t_1}}}{e^{x+W_{t_1}}+1}=\frac{1}{2}\int_{t_1}^{t_2}\mathsf{E}\,\frac{e^{x+W_t}-e^{2x+2W_t}}{(e^{x+W_t}+1)^3}\,dt.$$

Користуючись формулою (3) при $k=2$, отримаємо

$$L_0(x,t_2)-L_0(x,t_1)=\frac{1}{2}\int_{t_1}^{t_2}L_2(x,t)\,dt.$$

Отже,

$$\frac{\partial L_0(x,t)}{\partial t}=\frac{1}{2}L_2(x,t). \tag{4}$$

Ця формула без доведення та в інших позначеннях міститься в статті [5].

Диференціюючи рівність (4) по x, отримаємо

$$\frac{\partial L_k(x,t)}{\partial t}=\frac{1}{2}L_{k+2}(x,t),\qquad k\ge 0. \tag{5}$$

2.2. Нерівність, що пов'язує L_1, L_2 та L_3.

Лема 2.1. *Нехай ξ та η — незалежні випадкові величини, причому $\xi\sim N(0,1)$. Нехай $\zeta=\xi+\eta$. Позначимо щільність ζ через $p_\zeta(z)$. Тоді*

$$\frac{d^2}{dz^2}(\ln p_\zeta(z))=\operatorname{Var}[\eta\mid\zeta{=}z]-1, \tag{6}$$

де $\operatorname{Var}[\eta\mid\zeta{=}z]$ позначає умовну дисперсію η.

Доведення. Щільність ζ дорівнює

$$p_\zeta(z)=\mathsf{E}\,p_\xi(z-\eta)=\frac{1}{\sqrt{2\pi}}\,\mathsf{E}\,e^{-\frac{1}{2}(z-\eta)^2}.$$

Знайдемо похідні:

$$p'_\zeta(z) = \frac{1}{\sqrt{2\pi}} \mathsf{E}[(\eta - z)e^{-\frac{1}{2}(z-\eta)^2}],$$

$$\frac{d}{dz}(\ln p_\zeta(z)) = \frac{p'_\zeta(z)}{p_\zeta(z)} = \frac{\mathsf{E}[(\eta - z)e^{-\frac{1}{2}(z-\eta)^2}]}{\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}} = \frac{\mathsf{E}\,\eta e^{-\frac{1}{2}(z-\eta)^2}}{\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}} - z,$$

$$\begin{aligned}\frac{d^2}{dz^2}&(\ln p_\zeta(z)) \\ &= \frac{\mathsf{E}[\eta(\eta{-}z)e^{-\frac{1}{2}(z-\eta)^2}]\,\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2} - \mathsf{E}\,\eta e^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}[(\eta{-}z)e^{-\frac{1}{2}(z-\eta)^2}]}{\left(\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}\right)^2} - 1 \\ &= \frac{\mathsf{E}\,\eta^2 e^{-\frac{1}{2}(z-\eta)^2}\,\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2} - \left(\mathsf{E}\,\eta e^{-\frac{1}{2}(z-\eta)^2}\right)^2}{\left(\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}\right)^2} - 1.\end{aligned} \tag{7}$$

Якщо η має щільність $p_\eta(y)$, то умовний розподіл $[\eta \mid \zeta{=}z]$ має щільність

$$p_{\eta|\zeta=z}(y) = \frac{p_\eta(y)p_{\zeta|\eta=y}(z)}{\int p_\eta(y_1)p_{\zeta|\eta=y_1}(z)\,dy_1} = \frac{p_\eta(y)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-y)^2}}{\int p_\eta(y_1)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-y_1)^2}\,dy_1} = \frac{p_\eta(y)e^{-\frac{1}{2}(z-y)^2}}{\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}}.$$

Тому

$$\mathsf{E}[\eta \mid \zeta{=}z] = \frac{\mathsf{E}\,\eta e^{-\frac{1}{2}(z-\eta)^2}}{\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}}, \qquad \mathsf{E}[\eta^2 \mid \zeta{=}z] = \frac{\mathsf{E}\,\eta^2 e^{-\frac{1}{2}(z-\eta)^2}}{\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}}.$$

Наведені формули справедливі також у випадку, коли розподіл η не є абсолютно неперервним. Для доведення замість щільності $p_{\eta|\zeta=z}$ відносно міри Лебега можна використовувати щільність $\frac{dP_{\eta|\zeta=z}}{dP_\eta}$ відносно ймовірнісної міри маргінального розподілу η.

Отже,

$$\mathrm{Var}[\eta \mid \zeta{=}z] = \frac{\mathsf{E}\,\eta^2 e^{-\frac{1}{2}(z-\eta)^2}}{\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}} - \left(\frac{\mathsf{E}\,\eta e^{-\frac{1}{2}(z-\eta)^2}}{\mathsf{E}\, e^{-\frac{1}{2}(z-\eta)^2}}\right)^2. \tag{8}$$

З рівностей (7) та (8) отримаємо формулу (6). □

Наслідок 2.2. *Нехай ξ та η — незалежні випадкові величини, причому ξ має нормальний розподіл, $\xi \sim N(\mu, \sigma^2)$. Позначимо $\zeta = \xi + \eta$; $p_\zeta(z)$ — щільність ζ. Тоді*

$$\frac{d^2}{dz^2}(\ln p_\zeta(z)) = \frac{1}{\sigma^4}\mathrm{Var}[\eta \mid \zeta{=}z] - \frac{1}{\sigma^2}. \tag{9}$$

Зокрема,

$$\frac{d^2}{dz^2}(\ln p_\zeta(z)) \geq -\frac{1}{\sigma^2},$$

причому рівність досягається тоді і тільки тоді, коли розподіл η зосереджений в одній точці.

Приклад 2.1. Нехай $X \sim N(\mu_x, \sigma_x^2)$ та $U \sim N(0, \sigma_u^2)$ — дві незалежні нормально розподілені випадкові величини. Позначимо $W = X + U$, тоді $W \sim N(\mu_x, \sigma_w^2)$, де $\sigma_w^2 = \sigma_x^2 + \sigma_u^2$. Маємо

$$\forall w \in \mathbb{R}: \ \frac{d^2}{dw^2}(\ln p_W(w)) = -\frac{1}{\sigma_w^2},$$

$$\mathrm{Var}[X|W] = \frac{\sigma_u^2\sigma_x^2}{\sigma_w^2}$$

(ця формула наведена без доведення, наприклад, в [7]; загальне твердження доведено в [1, розділ 2.5]). Застосовуючи наслідок 2.2, бачимо, що (9) перетворюється у правильну рівність
$$-\frac{1}{\sigma_w^2} = \frac{1}{\sigma_u^4} \cdot \frac{\sigma_u^2 \sigma_x^2}{\sigma_w^2} - \frac{1}{\sigma_u^2}.$$

Приклад 2.2. Нагадаємо, що при фіксованому $\sigma^2 > 0$ функція $L_1(x, \sigma^2)$ є щільністю суми двох незалежних випадкових величин, одна з яких має логістичний розподіл, а інша — нормальний розподіл $N(0, \sigma^2)$. За наслідком 2.2
$$\frac{L_3(x,\sigma^2)L_1(x,\sigma^2) - L_2(x,\sigma^2)^2}{L_1(x,\sigma^2)^2} = \frac{\partial^2}{\partial x^2}(\ln L_1(x,\sigma^2)) > -\frac{1}{\sigma^2}. \tag{10}$$
З урахуванням того, що $L_1(x, \sigma^2) > 0$, нерівність (10) можна записати у вигляді
$$\left(L_3(x,\sigma^2) - \frac{L_2(x,\sigma^2)^2}{L_1(x,\sigma^2)}\right)\sigma^2 + L_1(x,\sigma^2) > 0, \qquad \sigma^2 \geq 0 \tag{11}$$
(на відміну від (10), нерівність (11) справедлива також при $\sigma^2 = 0$).

2.3. Оцінка знизу для $L_0(x, \sigma^2)$.

Доведемо нерівність
$$L_0(x,\sigma^2) \geq \frac{1}{2} e^{-|x|}. \tag{12}$$
При $x \geq 0$ маємо
$$L_0(x,\sigma^2) \geq \frac{1}{2} \geq \frac{1}{2} e^{-|x|},$$
і нерівність (12) правильна. При $x < 0$
$$\forall t{\geq}0 : \ \frac{d}{dt} L_0(x,t) = \frac{1}{2} L_2(x,t) > 0;$$
$$L_0(x,\sigma^2) \geq L_0(x,0) = \frac{e^x}{e^x+1} > \frac{1}{2} e^{-|x|},$$
і нерівність (12) справедлива також. Рівність в (12) досягається лише при $x = 0$.

2.4. Оцінки для $L_1(x, \sigma^2)$ та $L_2(x, \sigma^2)$.

Лема 2.3. *При всіх $x \in \mathbb{R}$ та $\sigma^2 \geq 0$ виконується нерівність*
$$|L_2(x,\sigma^2)| < L_1(x,\sigma^2) \leq L_0(x,\sigma^2)(1 - L_0(x,\sigma^2)).$$
Рівність досягається при $\sigma^2 = 0$.

Доведення. Безпосередньо доводиться нерівність
$$\frac{e^\xi}{(e^\xi+1)^2} + \frac{e^\eta}{(e^\eta+1)^2} \leq \frac{e^\xi}{(e^\xi+1)(e^\eta+1)} + \frac{e^\eta}{(e^\xi+1)(e^\eta+1)},$$
причому рівність досягається лише при $\xi = \eta$. Візьмемо математичне сподівання від обох частин нерівності, вважаючи що ξ та η — незалежні нормально розподілені випадкові величини, $\xi \sim \eta \sim N(x, \sigma^2)$:
$$L_1(x,\sigma^2) + L_1(x,\sigma^2) \leq L_0(x,\sigma^2)(1 - L_0(x,\sigma^2)) + (1 - L_0(x,\sigma^2))L_0(x,\sigma^2),$$
$$L_1(x,\sigma^2) \leq L_0(x,\sigma^2)(1 - L_0(x,\sigma^2)).$$
Рівність досягається, якщо $\xi = \eta$ майже напевно, тобто при $\sigma^2 = 0$.

Нерівність $|L_2(x,\sigma^2)| < L_1(x,\sigma^2)$ отримаємо, беручи математичне сподівання в правильній нерівності
$$-\frac{e^\xi}{(1+e^\xi)^2} < \frac{e^\xi - e^{2\xi}}{(1+e^\xi)^3} < \frac{e^\xi}{(1+e^\xi)^2},$$
де $\xi \sim N(x, \sigma^2)$. □

3. Оцінка максимальної вірогідності (ОМВ)

Розглянемо регресію Y за X^{obs}:

$$\begin{aligned}\mathsf{P}[Y{=}1 \mid X^{\mathrm{obs}}] &= \int_{-\infty}^{\infty} \mathsf{P}\left[Y{=}1 \mid X^{\mathrm{tr}}{=}x, X^{\mathrm{obs}}\right] \, p_{X^{\mathrm{tr}}|X^{\mathrm{obs}}}(x)\, dx \\ &= \mathsf{E}\left[\frac{e^{\beta_0+\beta_1 X^{\mathrm{tr}}}}{e^{\beta_0+\beta_1 X^{\mathrm{tr}}}+1} \,\middle|\, X^{\mathrm{obs}}\right] = L_0\left(\beta_0+\beta_1 X^{\mathrm{obs}}, \beta_1^2\tau^2\right).\end{aligned}$$

Оцінка максимальної вірогідності знаходиться як точка максимуму функціоналу на параметричній множині Θ:

$$\sum_{Y_n=0} \ln\left(1 - L_0\left(b_0+\beta_1 X_n^{\mathrm{obs}}, b_1^2\tau^2\right)\right) + \sum_{Y_n=1} \ln L_0\left(b_0+\beta_1 X_n^{\mathrm{obs}}, b_1^2\tau^2\right) \to \max, \tag{13}$$

або

$$\sum_{n=1}^{N} \ln L_0\left((2Y_n-1)\left(b_0+b_1 X_n^{\mathrm{obs}}\right), b_1^2\tau^2\right) \to \max,$$

де N — кількість спостережень, $b=(b_0,b_1)$ — змінна, за якою максимізуємо функціонал.

Диференціюючи (13) та зводячи до спільного знаменника, отримаємо систему рівнянь для оцінки (якщо вона буде внутрішньою точкою параметричної множини):

$$\sum_{n=1}^{N} \frac{L_0(\hat\beta_0+\hat\beta_1 X_n^{\mathrm{obs}}, \hat\beta_1^2\tau^2) - Y_n}{L_0(\hat\beta_0+\hat\beta_1 X_n^{\mathrm{obs}}, \hat\beta_1^2\tau^2)(1-L_0(\hat\beta_0+\hat\beta_1 X_n^{\mathrm{obs}}, \hat\beta_1^2\tau^2))} \frac{\partial L_0(\hat\beta_0+\hat\beta_1 X_n^{\mathrm{obs}}, \hat\beta_1^2\tau^2)}{\partial\hat\beta} = 0,$$

тобто

$$\begin{cases} \displaystyle\sum_{n=1}^{N} \frac{(L_0-Y_n)\cdot L_1}{L_0\cdot(1-L_0)} = 0, \\ \displaystyle\sum_{n=1}^{N} \frac{(L_0-Y_n)(L_1 X_n^{\mathrm{obs}} + \hat\beta_1\tau^2 L_2)}{L_0\cdot(1-L_0)} = 0, \end{cases} \tag{14}$$

де L_0, L_1, L_2 беруться в точці $(\hat\beta_0+\hat\beta_1 X_n^{\mathrm{obs}}, \hat\beta_1^2\tau^2)$. Отже, значення виразів, які тут позначено через через L_0, L_1 та L_2, різні для різних доданків та залежать від невідомих змінних системи рівнянь.

4. Єдиність розв'язку граничного рівняння

4.1. Дослідження рівняння $\boldsymbol{L_0(\beta_0+\beta_1 x, \beta_1^2\tau^2) = L_0(\gamma_0+\gamma_1 x, \gamma_1^2\tau^2)}$.

Лема 4.1. *Якщо $\beta_0\ne\gamma_0$ або $\beta_1\ne\gamma_1$, то рівняння*

$$L_0(\beta_0+\beta_1 x, \beta_1^2\tau^2) = L_0(\gamma_0+\gamma_1 x, \gamma_1^2\tau^2) \tag{15}$$

має не більше одного розв'язку.

Доведення. Якщо $\beta_1=\gamma_1$, то $\beta_0\ne\gamma_0$, і для всіх x справедлива нерівність $\beta_0+\beta_1 x \ne \gamma_0+\gamma_1 x$, проте $\beta_1^2\tau^2 = \gamma_1^2\tau^2$. Оскільки при фіксованому σ^2 функція $L_1(z,\sigma^2)$ зростає по z, то рівняння (15) не має розв'язків.

Тепер розглянемо випадок $\beta_1\ne\gamma_1$. Не втрачаючи загальності, вважатимемо, що $\beta_1>\gamma_1$. Припустимо, що рівняння (15) має розв'язок x_1. Доведемо нерівність

$$\frac{d}{dx}L_0(\beta_0+\beta_1 x, \beta_1^2\tau^2) > \frac{d}{dx}L_0(\gamma_0+\gamma_1 x, \gamma_1^2\tau^2) \tag{16}$$

в точці $x=x_1$.

Функція $z \mapsto L_0(z, \sigma^2)$ при фіксованому σ^2 є бієкцією $\mathbb{R} \to (0,1)$. Означимо $\mu(\tilde{b}_1)$ з рівняння

$$L_0(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2) = L_0(\beta_0 + \beta_1 x_1, \beta_1^2\tau^2).$$

Зауважимо, що $\mu(\beta_1) = \beta_0 + \beta_1 x_1$ та $\mu(\gamma_1) = \gamma_0 + \gamma_1 x_1$. За теоремою про неявну функцію існує похідна $\mu'(\tilde{b}_1)$, знайдемо її з рівняння

$$\begin{gathered}\frac{\partial}{\partial \tilde{b}_1} L_0(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2) = 0;\\ L_1(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2) \cdot \mu'(\tilde{b}_1) + \tilde{b}_1\tau^2 L_2(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2) = 0,\\ \mu'(\tilde{b}_1) = -\frac{\tilde{b}_1\tau^2 L_2(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2)}{L_1(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2)}.\end{gathered}$$

Тепер знайдемо знак похідної:

$$\begin{aligned}&\frac{\partial}{\partial \tilde{b}_1}(\tilde{b}_1 L_1(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2))\\ &\qquad = L_1(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2) + \tilde{b}_1 L_2(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2)\mu'(\tilde{b}_1) + \tilde{b}_1^2\tau^2 L_3(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2)\\ &\qquad = L_1 - \frac{\tilde{b}_1^2\tau^2 L_2^2}{L_1} + \tilde{b}_1^2\tau^2 L_3 > 0\end{aligned}$$

за нерівністю (10), записаною у вигляді $\frac{L_3 L_1 - L_2^2}{L_1^2} > -\frac{1}{\tilde{b}_1^2\tau^2}$. Тут L_1, L_2, L_3 беруться в точці $(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2)$. Отже, функція $\tilde{b}_1 L_1(\mu(\tilde{b}_1), \tilde{b}_1^2\tau^2)$ зростає по $\tilde{b}_1$. Нагадаємо припущення $\beta_1 > \gamma_1$. Маємо

$$\begin{gathered}\beta_1 L_1(\mu(\beta_1), \beta_1^2\tau^2) > \gamma_1 L_1(\mu(\gamma_1), \gamma_1^2\tau^2),\\ \beta_1 L_1(\beta_0 + \beta_1 x_1, \beta_1^2\tau^2) > \gamma_1 L_1(\gamma_0 + \gamma_1 x_1, \gamma_1^2\tau^2),\\ \frac{\partial}{\partial x}(L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2))\big|_{x=x_1} > \frac{\partial}{\partial x}(L_0(\gamma_0 + \gamma_1 x, \gamma_1^2\tau^2))\big|_{x=x_1}.\end{gathered}$$

Ми довели нерівність (16).

Отже, неперервна функція $L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2) - L_0(\gamma_0 + \gamma_1 x, \gamma_1^2\tau^2)$ в тих точках, де вона дорівнює 0, має додатну похідну. Така функція може дорівнювати 0 не більш як у одній точці. □

Зауваження 4.1. З доведення леми випливає, що при $\beta \neq \gamma$ функція $L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2) - L_0(\gamma_0 + \gamma_1 x, \gamma_1^2\tau^2)$ змінює знак у тих точках, де вона дорівнює 0.

Окрім того, якщо рівняння (15) має розв'язок $x = x_1$, то

$$\operatorname{sign}(L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2) - L_0(\gamma_0 + \gamma_1 x, \gamma_1^2\tau^2)) = \operatorname{sign}(\beta_1 - \gamma_1)\operatorname{sign}(x - x_1), \tag{17}$$

де $\operatorname{sign}(x) = -1$ при $x < 0$, $\operatorname{sign}(x) = 0$ при $x = 0$, та $\operatorname{sign}(x) = 1$ при $x > 0$.

Якщо $\beta \neq \gamma$, то рівність (17) випливає з доведення леми 4.1. Якщо ж $\beta = \gamma$, то рівність (17) тривіальна.

4.2. **Умовне математичне сподівання оціночної функції.** Систему рівнянь для оцінки параметра регресії методом максимальної вірогідності (14) запишемо так:

$$\sum_{n=1}^{N} s(X_n^{\text{obs}}, Y_n; b) = 0,$$

де $s(X^{\text{obs}}, Y; b)$ — елементарна оціночна функція,

$$s(x, y; b) = \begin{pmatrix} \dfrac{(L_0 - y)L_1}{L_0 \cdot (1 - L_0)} \\ \dfrac{(L_0 - y)(xL_1 + b_1\tau^2 L_2)}{L_0 \cdot (1 - L_0)} \end{pmatrix},$$

де L_0, L_1, L_2 беруться в точці $(b_0 + b_1 x, b_1^2\tau^2)$.

Будемо досліджувати умовне математичне сподівання

$$s_0(X^{\text{obs}}; b) = \mathsf{E}[s(X^{\text{obs}}, Y; b) \mid X^{\text{obs}}],$$

$$s_0(x; b) = \frac{L_0(b_0 + b_1 x, b_1^2\tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2)}{L_0(b_0 + b_1 x, b_1^2\tau^2)(1 - L_0(b_0 + b_1 x, b_1^2\tau^2))} \begin{pmatrix} L_1 \\ xL_1 + b_1\tau^2 L_2 \end{pmatrix},$$

де L_1 та L_2 знову беруться в точці $(b_0+b_1 x, b_1^2\tau^2)$. У цьому розділі β позначає істинне значення параметра регресії. (Залежність $s_0(x; b)$ від істинного параметра регресії β ніяк позначати не будемо, а будемо вважати β фіксованим). Очевидно, що

$$\forall x \in \mathbb{R}: \; s_0(x; \beta) = 0.$$

Лема 4.2. *Для кожного $b \neq \beta$ ($b \in \mathbb{R}^2$) існує множник $\theta \in \mathbb{R}^2$, такий що для всіх $x_0 \in \mathbb{R}$ має місце нерівність*

$$\theta^\top s_0(x_0, b) \geq 0,$$

причому рівність досягається не більш як при одному x_0.

Запишемо твердження леми за допомогою кванторів:

$$\forall b \neq \beta \; \exists \theta \in \mathbb{R}^2 \; \exists x_1 \in \mathbb{R}:$$

$$\begin{cases} \forall x_0 \in \mathbb{R}: \; \theta^\top s_0(x_0, b) \geq 0, \\ \forall x_0 \neq x_1: \; \theta^\top s_0(x_0, b) > 0. \end{cases} \tag{18}$$

Доведення. Зафіксуємо $b \neq \beta$. Рівняння

$$L_0(b_0 + b_1 x, b_1^2\tau^2) = L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2) \tag{19}$$

може або не мати розв'язків, або мати лише один розв'язок.

Спочатку розглянемо випадок, коли рівняння не має розв'язків. Тоді вираз $L_0(b_0+b_1 x, b_1^2\tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2)$ набуває значень лише одного знаку. Покладемо $\theta = (1, 0)^\top$, якщо цей вираз набуває лише додатних значень, та $\theta = (-1, 0)^\top$, якщо цей вираз набуває лише від'ємних значень. З урахуванням того, що

$$\frac{L_1(b_0 + b_1 x, b_1^2\tau^2)}{L_0(b_0 + b_1 x, b_1^2\tau^2)(1 - L_0(b_0 + b_1 x, b_1^2\tau^2))} > 0$$

(див. розділ 2.1), твердження леми виконується.

Тепер розглянемо випадок, коли рівняння (19) має один розв'язок. Цей розв'язок і візьмемо за x_1. Згідно із зауваженням 4.1, для всіх $x \in \mathbb{R}$

$$\operatorname{sign}(L_0(b_0 + b_1 x, b_1^2\tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2\tau^2)) = \operatorname{sign}(b_1 - \beta_1) \operatorname{sign}(x - x_1), \tag{20}$$

причому $b_1 \neq \beta_1$.

Покладемо $\theta = (\theta_1, \theta_2)^\top$, де

$$\theta_2 = \operatorname{sign}(b_1 - \beta_1),$$

$$\theta_1 = \left(-x_1 - b_1\tau^2 \frac{L_2(b_0 + b_1 x_1, b_1^2\tau^2)}{L_1(b_0 + b_1 x_1, b_1^2\tau^2)}\right) \theta_2.$$

Розглянемо функцію

$$f(x) = \theta_1 L_1(b_0 + b_1 x, b_1^2\tau^2) + \theta_2(xL_1(b_0 + b_1 x, b_1^2\tau^2) + b_1\tau^2 L_2(b_0 + b_1 x, b_1^2\tau^2)).$$

Маємо $f(x_1) = 0$ та

$$f'(x) = \theta_1 b_1 L_2 + \theta_2 L_1 + \theta_2 b_1 x L_2 + \theta_2 b_1^2 \tau^2 L_3, \tag{21}$$

де L_1 та L_2 беруться в точці $(b_0 + b_1 x, b_1^2 \tau^2)$.

Доведемо, що

$$\forall x : \quad \bigl(f(x) = 0 \Rightarrow \operatorname{sign}(f'(x)) = \operatorname{sign}(b_1 - \beta_1)\bigr) \tag{22}$$

(тобто функція $f(x)$ в точках, де вона дорівнює нулю, має похідну одного знаку).

Справді, якщо $f(x) = 0$, то

$$\theta_1 = -\frac{x L_1 + b_1 \tau^2 L_2}{L_1} \theta_2.$$

Підставимо значення θ_1 у вираз для $f'(x)$

$$\begin{aligned} f'(x) &= -\theta_2 \cdot \frac{x L_1 + b_1 \tau^2 L_2}{L_1} \cdot b_1 L_2 + \theta_2 L_1 + \theta_2 b_1 x L_2 + \theta_2 b_1^2 \tau^2 L_3 \\ &= \left(-\beta_1^2 \tau^2 \frac{L_2^2}{L_1} + L_1 + b_1^2 \tau^2 \right) \theta_2, \end{aligned}$$

і за нерівностями $\frac{L_3 L_1 - L_2^2}{L_1^2} > -\frac{1}{b_1^2 \tau^2}$ (це — нерівність (10)) та $L_1 > 0$ отримаємо формулу з (22).

Неперервна функція, що задовольняє (22), може досягати 0 не більш як в одній точці. З урахуванням $f(x_1) = 0$ отримаємо, що

$$\operatorname{sign}(f(x)) = \operatorname{sign}(b_1 - \beta_1) \operatorname{sign}(x - x_1). \tag{23}$$

Оскільки

$$\theta^\top s_0(x; b) = \frac{L_0(b_0 + b_1 x, b_1^2 \tau^2) - L_0(\beta_0 + \beta_1 x, \beta_1^2 \tau^2)}{L_0(b_0 + b_1 x, b_1^2 \tau^2)(1 - L_0(b_0 + b_1 x, b_1^2 \tau^2))} \cdot f(x),$$

то з співвідношень (20), (23), $0 < L_0(b_1 + b_1 x, b_1^2 \tau^2) < 1$ та $b_1 \neq \beta_1$ випливає рівність

$$\operatorname{sign}(\theta^\top s_0(x; b)) = \operatorname{sign}(x - x_1)^2,$$

з якої випливає система нерівностей (18). □

4.3. **Єдиність розв'язку граничного оціночного рівняння в структурній моделі.** Позначимо граничну оціночну функцію

$$s_\infty(b) = \mathsf{E}\, s(X^{\text{obs}}, Y; b), \tag{24}$$

якщо математичне сподівання існує. Тут $s(X^{\text{obs}}, Y; b)$ — елементарна оціночна функція, означена в розділі 4.2; сумісний розподіл X^{obs} та Y задовольняє (1)–(2). Функція $s_\infty(b)$ залежить також від істинного значення параметра β та розподілу X^{obs}, але ми вважаємо цей розподіл та β фіксованими.

Лема 4.3. *Якщо розподіл X^{obs} такий, що $\mathsf{E}\,|X^{\text{obs}}| < \infty$, то математичне сподівання в (24) скінченне.*

Доведення. Згідно з лемою 2.3 координати вектора $s(X^{\text{obs}}, Y; b)$ обмежені інтегровними випадковими величинами:

$$|s^{(1)}(X^{\text{obs}}, Y; b)| \le |L_0(b_0 + b_1 X^{\text{obs}}, b_1^2 \tau^2) - Y| < 1, \tag{25}$$

$$\begin{aligned} \left| s^{(2)} \left(X^{\text{obs}}, Y; b\right) \right| &\le \left| L_0 \left(b_0 + b_1 X^{\text{obs}}, b_1^2 \tau^2\right) - Y \right| \cdot (|X^{\text{obs}}| + |b_1| \tau^2) \\ &\le |X^{\text{obs}}| + |b_1| \tau^2. \end{aligned} \tag{26}$$

Тому вектор $s(X^{\text{obs}}, Y; b)$ має скінченне математичне сподівання. □

Якщо математичне сподівання $s(X^{\mathrm{obs}}, Y; b)$ існує, то воно дорівнює

$$\mathsf{E}\, s(X^{\mathrm{obs}}, Y; b) = \mathsf{E}[\mathsf{E}[s(X^{\mathrm{obs}}, Y; b) \mid X^{\mathrm{obs}}]] = \mathsf{E}\, s_0(X^{\mathrm{obs}}; b). \tag{27}$$

Твердження 4.4. *Якщо розподіл X^{obs} не зосереджений в одній точці та математичне сподівання $s(X^{\mathrm{obs}}, Y; b)$ скінченне, то*

$$s_\infty(b) = 0 \Leftrightarrow b = \beta.$$

Доведення. Якщо математичне сподівання вектора $s(X^{\mathrm{obs}}, Y; \beta)$ скінченне, то за рівністю (27) виконується $\mathsf{E}\, s(X^{\mathrm{obs}}, Y; \beta) = 0$, бо $s_0(X^{\mathrm{obs}}; \beta) = 0$.

Якщо $b \neq \beta$, то за лемою 4.2 існують такі $\theta \in \mathbb{R}^2$ та $x_1 \in \mathbb{R}$, що

$$\theta^\top s_0(X^{\mathrm{obs}}; b) \geq 0,$$
$$\theta^\top s_0(X^{\mathrm{obs}}; b) > 0 \quad \text{при} \quad X^{\mathrm{obs}} \neq x_1.$$

Тоді, за формулою (27) отримаємо

$$\theta^\top s_\infty(b) = \mathsf{E}[\theta^\top s_0(X^{\mathrm{obs}}; b)] > 0,$$

враховуючи, що розподіл X^{obs} не зосереджений в точці x_1. Тому $s(b) \neq 0$. □

Наслідок 4.5. *Якщо $\mathsf{E}\,|X^{\mathrm{obs}}| < \infty$ та розподіл X^{obs} не зосереджений в одній точці, то рівняння $s_\infty(b) = 0$ має рівно один розв'язок, а саме $b = \beta$.*

5. Консистентність оцінки максимальної вірогідності

Для того, щоб скористатись теоремою про консистентність ОМВ для векторного параметра, штучно припустимо, що розподіл $P_{X^{\mathrm{obs}}}$ спостережуваного сурогатного регресора X^{obs} відомий. Знання чи незнання цього розподілу не впливає на оцінку максимальної вірогідності параметра β. За мажорантну міру для розподілу пари (X^{obs}, Y) візьмемо добуток мір $\nu = P_{X^{\mathrm{obs}}} \otimes (\delta_0 + \delta_1)$. Тут δ_0 та δ_1 — вироджені ймовірнісні міри на $\{0, 1\}$, зосереджені в точках 0 та 1. Це дозволяє вилучити множник $p_{X^{\mathrm{obs}}}(x)$ з функції вірогідності.

Твердження 5.1. *Нехай в структурній моделі параметрична множина $\Theta \subset \mathbb{R}^2$ замкнена, обмежена, містить істинне значення параметра β як внутрішню точку. Нехай*

$$\hat\beta_{\mathrm{ML}} = \operatorname*{arg\,max}_{b \in \Theta} \sum_{n=1}^{N} L_0((2Y_n - 1)\,(b_0 + b_1 X_n^{\mathrm{obs}}), b_1^2 \tau^2)$$

є оцінкою максимальної вірогідності. Якщо розподіл X^{obs} не зосереджений в одній точці та $\mathsf{E}\,|X^{\mathrm{obs}}| < \infty$, то $\hat\beta_{\mathrm{ML}}$ — сильно консистентна оцінка параметра β, тобто майже напевно $\hat\beta_{\mathrm{ML}} \to \beta$ при $N \to \infty$.

Доведення. Скористаємось теоремою про консистентність ОМВ з підручника [6]. Функція вірогідності для одного спостереження дорівнює

$$f(x, y; \beta) = L_0((2y - 1)\,(\beta_0 + \beta_1 x), \beta_1^2 \tau^2).$$

Перевіримо умови консистентності:
(c1) Модель ідентифіковна, тобто для всіх $b \neq \beta$

$$\nu(\{(x, y) \mid f(x, y; b) \neq f(x, y; b)\}) > 0.$$

Ця умова рівносильна нерівності

$$\mathsf{P}(L_0(b_0 + b_1 X^{\mathrm{obs}}, b_1^2 \tau^2) \neq L_0(\beta_0 + \beta_1 X^{\mathrm{obs}}, \beta_1^2 \tau^2)) > 0, \qquad b \neq \beta,$$

та виконується згідно з лемою 4.1.
(c2) Параметрична множина Θ є компактом в $\mathbb{R}^2$, який містить істинне значення параметра β як внутрішню точку.

(c3) При всіх (x, y) функція вірогідності $f(x, y; b)$ строго додатна та неперервно диференційовна за b, причому випадкові величини $\ln f(X^{\rm obs}, Y; b)$ та $\frac{\partial}{\partial b}(\ln f(X^{\rm obs}, Y; b))$ мажоруються інтегровною випадковою величиною.

Мажоровність $\ln f(X^{\rm obs}, Y; b)$ випливає з нерівності (12):

$$-|b_0 + b_1 X^{\rm obs}| - \ln 2 \le \ln f(X^{\rm obs}, Y; b) < 0.$$

Мажоровність $\frac{\partial}{\partial b}(\ln f(X^{\rm obs}, Y; b)) = -s(X^{\rm obs}, Y; b)$ має місце згідно нерівностей (25)–(26). □

6. Висновки

У роботі доведено сильну консистентність оцінки максимальної вірогідності у структурній моделі Берксона логістичної регресії.

Розглянемо структурну модель логістичної регресії з одним нормально розподіленим регресором з класичною похибкою. Шляхом калібрування регресії, тобто шляхом лінійної заміни незалежної змінної, ця модель зводиться до моделі Берксона [3, розділ 2.2.3]. Застосовуючи твердження 4.4 для каліброваної моделі, можна перевірити, що логістична регресія з класичною похибкою задовольняє умову (vii) статті [8].

Розглянуті мною, але не включені в цю статтю, окремі випадки дозволяють висунути наступну гіпотезу.

Гіпотеза. *Якщо не всі $X_n^{\rm obs}$ однакові, то система рівнянь* (14) *для оцінки максимальної вірогідності має не більше одного розв'язку на $\mathbb{R}^2$.*

Якщо гіпотеза виявиться справедливою, то це дозволить спростити побудову оцінки максимальної вірогідності. По-перше, не буде потрібно вимагати, щоб була задана обмежена параметрична множина Θ. По-друге, стаціонарна точка функції вірогідності автоматично буде точкою глобального максимуму.

Література

1. Т. Андерсон, *Введение в многомерный статистический анализ*, “Физматгиз”, Москва, 1963.
2. D. Burr, *On errors-in-variables in binary regression — Berkson case*, Journal of the American Statistical Association **83** (1988), no. 403, 739–743.
3. R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, Second edition, Chapman & Hall, London, 2006.
4. R. J. Carroll, C. H. Spiegelman, K. K. G. Lan, K. T. Bailey, and R. D. Abbott, *On errors in variables in binary regression models*, Biometrika **71** (1984), 19–26.
5. E. Crouch and D. Spiegelman, *The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)\,dt$: application to logistic-normal models*, Journal of the American Statistical Association **85** (1990), no. 410, 464–469.
6. М. В. Карташов, *Імовірність. Процеси. Статистика*, Київський університет, Київ, 2008.
7. H. Küchenhoff, *The identification of logistic regression models with errors in the variables*, Statistical Papers **36** (1995), 41–47.
8. A. Kukush and H. Schneeweiss, *Comparing different estimators in a nonlinear measurement error model. Part I*, Mathematical Methods of Statistics **14** (2005), 53–79.

Кафедра теорії ймовірностей, математичної статистики та актуарної математики, механіко-математичний факультет, Київський національний університет імені Тараса Шевченка; Володимирська, 64; Київ-33; 01601 Україна

Адреса електронної пошти: shklyar@mail.univ.kiev.ua

Надійшла 07/09/2011