УДК 519.21

Лукашевич С. О., Ямненко Р. Є.

https://doi.org/10.18523/2617-70806202324-27

ПОХІДНІ ФУНКЦІЇ ВІРОГІДНОСТІ ДЛЯ ЛІНІЙНОЇ ЗМІШАНОЇ МОДЕЛІ З ПРИПУЩЕННЯМ ПРО СКЛАДНУ СИМЕТРІЮ

У роботі досліджено властивості лінійних змішаних моделей із простими випадковими ефектами виду

$$y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i, i = 1, \ldots, M, \quad \gamma_i \sim N(0, \Psi), \varepsilon_i \sim N(0, \sigma^2 I),$$

де β – p-вимірний вектор фіксованих ефектів, γ_i – q-вимірний вектор випадкових ефектів, X_i та Z_i – відомі матриці регресорів розмірностей $n_i \times p$ і $n_i \times q$, а ε_i – вектори внутрішньогрупових похибок зі сферичним гауссівським розподілом. Припускаючи складну симетрію кореляційної структури залежності між внутрішньогруповими похибками, отримано аналітичні формули для перших двох часткових похідних за кореляційними параметрами моделі.

Ключові слова: змішана лінійна модель, обмежена оцінка максимальної вірогідності, похідна, випадкові ефекти, складна симетрія.

Вступ

Припустимо, що спостерігаються дані з n точок $y_1, \ldots, y_n$, які ми хочемо пояснити за допомогою n значень для кожної з p пояснювальних змінних $x_{11}, \ldots, x_{n1}$, $x_{21}, \ldots x_{2n}, \ldots, x_{p1}, \ldots, x_{pn}$. Значення x_{ij} можуть бути як неперервними регресійними змінними типу регресії, так і фіктивними змінними, що вказують на належність до класу і набувають значення 0 чи 1. Стандартна лінійна модель для такої задачі має вигляд (див. наприклад, [1])

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_n, \quad i = 1, \ldots, n, \qquad (1)$$

де β_j – це невідомі коефіцієнти фіксованих ефектів, а $\varepsilon_1, \ldots, \varepsilon_n$ – невідомі незалежні між собою випадкові похибки з математичним сподіванням 0 і однаковою дисперсією σ^2. Втім, зручніше записувати модель (1) у матричній формі

$$y = X\beta + \varepsilon, \qquad (2)$$

де y – це вектор спостережених y_i, X – відома матриця з x_{ij}, β – невідомий вектор параметрів, ε – випадковий вектор незалежних однаково розподілених гауссівських похибок. На жаль, у багатьох реальних ситуаціях таке припущення щодо ε не відповідає дійсності, де стають у нагоді змішані моделі, що дозволяють гнучке задання коваріаційної матриці випадкового вектора ϵ, тобто такі моделі можуть мати і корельовані похибки з неоднорідними дисперсіями.

Лінійні змішані моделі можна записати у вигляді

$$y = X\beta + Z\gamma + \varepsilon, \qquad (3)$$

де додатково до моделі (2) з'являється вектор невідомих параметрів γ для випадкових ефектів, Z – відома матриця розмірності $n \times q$. Будемо також припускати, що $\mathsf{E}(\gamma) = 0, \mathsf{E}(\varepsilon) = 0$ та $\mathsf{Cov}(\gamma, \varepsilon) = 0$. Нехай $G = \mathsf{Var}(\gamma), R = \mathsf{Var}(\varepsilon)$, тоді коваріаційна матриця вектора спостережень y дорівнює

$$V = ZGZ' + R. \qquad (4)$$

Отже, можна моделювати коваріаційну структуру y, задаючи модельну матрицю випадкових ефектів Z і вказуючи коваріаційні структури для G і R. Таку модель називають змішаною, бо вона пояснює вплив одночасно як фіксованих, так і випадкових ефектів.

У разі простих однорівневих випадкових ефектів задання моделі спрощується: Z складається з фіктивних бінарних змінних, G містить дисперсійні компоненти на головній діагоналі, а $R = \sigma^2 I_n$, I_n – одинична матриця.

Оцінювання параметрів змішаної моделі

Оцінювання параметрів змішаної моделі складніше, ніж загальної лінійної моделі, бо на додаток до невідомого вектора параметрів β фіксованих ефектів із лінійної моделі ще треба оцінити параметри γ, G та R. Метод найменших квадратів уже не підходить. Якби коваріаційна матриця V була відомою, можна було б застосувати узагальнений метод найменших квадратів

(GLS), який мінімізує значення функціоналу

$$(y - X\beta)'V^{-1}(y - X\beta).$$

Одначе, V ми не знаємо, як, відповідно, і матриці G і R. Будемо вважати, що випадкові ефекти мають нормальний розподіл $\gamma \sim N(0, G)$, $\varepsilon \sim N(0, R)$, $\mathsf{Cov}(\gamma, \varepsilon) = 0$. Метод максимальної вірогідності використовує припущення про нормальність розподілів γ і ε. Функція вірогідності L для моделі (3) – це щільність f для даних за умови фіксованих параметрів, але яка розглядається навпаки як функція від параметрів за умови фіксованих даних, тобто

$$L(\beta, \theta, \sigma^2|y) = f(y|\beta, \theta, \sigma^2),$$

де θ – вектор параметрів, що відповідає невідомим компонентам матриці G. Відповідна логарифмічна функція вірогідності має вигляд

$$l(\beta, \theta, \sigma^2|y) = -\frac{1}{2}\log|V| - \frac{1}{2}P'V^{-1}P - \frac{n}{2}\log(2\pi), \tag{5}$$

де $P = y - X(V^{-1}X)^{-1}X'V^{-1}y$.

Оцінки максимальної вірогідності дисперсійних компонент мають тенденцію недооцінювати ці параметри. Натомість рекомендують використовувати метод обмеженої максимальної вірогідності (REML):

$$L_{REML}(\theta, \sigma^2|y) = \int L(\beta, \theta, \sigma^2|y)\mathsf{d}\beta,$$

що в рамках баєсівської інтерпретації відповідає припущенню про локально рівномірний апріорний розподіл для фіксованих ефектів β, а за допомогою інтегрування можна від них позбавитися. Відповідна логарифмічна функція обмеженої вірогідності має вигляд

$$l_{REML}(\theta, \sigma^2|y) = -\frac{1}{2}\log|V| - \frac{1}{2}P'V^{-1}P - \tag{6}$$
$$-\frac{1}{2}\log|X'V^{-1}X| - \frac{n-p}{2}\log(2\pi),$$

де p – число фіксованих параметрів. Більш детально про теорію лінійних моделей можна прочитати у роботах [2; 4–6].

Мінімізація логарифмічної вірогідності

Популярними числовими методами оптимізації для розв'язання проблем мінімізації функцій є методи Ньютона–Рафсона і Левенберга–Марквардта [3]. Останній базується на методі Гаусса–Ньютона з використанням регуляризації Левенберга. У цьому методі параметри моделі оновлюються за допомогою вагового коефіцієнта, який регулюється в процесі ітерацій. Ефективне застосування цих та інших методів потребує знання аналітичного вигляду перших двох похідних від цільової функції. Зосередимося на їх знаходженні.

Залишкову дисперсію σ^2 можна аналітично виразити через інші параметри. Тоді, використовуючи профілювання вірогідності, тобто підставляючи оцінку $\hat{\sigma}^2$ у функцію вірогідності, можна зменшити число параметрів і покращити властивості збіжності розв'язку.

$$\widehat{\sigma^2} = \frac{y'\left(V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}\right)y}{n-p}. \tag{7}$$

Оцінки фіксованих параметрів дорівнюють

$$\hat{\beta} = (X'V^{-1}X)^{-1}V^{-1}y. \tag{8}$$

Складна симметрія

Для моделювання залежності між спостереженнями використовують різні кореляційні структури. Для змішаних лінійних моделей вони використовуються для моделювання залежності між внутрішньогруповими похибками. Складна симетрія – це найпростіша структура серійної кореляції, яка передбачає однакову кореляцію між усіма внутрішньогруповими похибками, що належать до однієї групи. Відповідна кореляційна модель має вигляд

$$\mathsf{cor}(\varepsilon_{ij}, \varepsilon_{ik}) = \rho, \forall j \neq k,$$

де єдиний кореляційний параметр ρ відомий також як коефіцієнт внутрішньокласової кореляції.

Коваріаційна матриця для i-го відгуку лінійної змішаної моделі з незалежними і однаково розподіленими внутрішньогруповими похибками з дисперсією σ^2 і єдиним простим випадковим ефектом із дисперсією σ_b^2 має вигляд $\sigma^2 I + \sigma^2 11'$. Відповідна кореляційна матриця дорівнює $\sigma^2/(\sigma^2 + \sigma_b^2)I + \sigma_b^2/(\sigma^2 + \sigma_b^2)11'$, тобто коефіцієнт внутрішньокласової кореляції $\rho = \sigma_b^2/(\sigma^2 + \sigma_b^2)$. Тут 1 – це матриця, кожен елемент якої дорівнює одиниці.

Отже, перепишемо модель (3) із одним рівнем групування так:

$$y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i, i = 1, \ldots, M, \tag{9}$$
$$\gamma_i \sim N(0, \Psi), \varepsilon_i \sim N(0, \sigma^2 I),$$

де M – число різних груп по n_i спостережень, $n = \sum_{i=1}^{M} n_i$. Випадкові ефекти γ_i та внутрішньо-групові похибки ε_i є незалежними для різних груп та між собою в одній групі. Коваріаційна матриця $V = \sum_{i=1}^{M} Z_i G_j Z_i' + \sigma^2 I$ має блочно-діагональну структуру. Нехай вектор параметрів θ складається з дисперсійних компонент матриці $\Theta = \frac{1}{\sigma^2}\Psi$. Розглянемо окремо кожен блок. Тоді

$$\frac{\partial V_i}{\partial\theta_j} = Z_i \frac{\partial\Theta}{\partial\theta_j} Z_i',$$

де $\frac{\partial\Theta}{\partial\theta_j} = \delta_{ij}$ – символ Кронекера. Позначимо

$$T_j = \frac{1}{\widehat{\sigma^2}}\left(y'\frac{\partial V_i}{\partial\theta_j}y - 2y'\frac{\partial V^{-1}}{\partial\theta_j}X\hat\beta + X'\frac{\partial V^{-1}}{\partial\theta_j}X\hat\beta\right),$$

де оцінки $\widehat{\sigma^2}$ і $\hat\beta$ визначено у (7) і (8) відповідно. Тоді формули для часткових похідних профільованої функції обмеженої вірогідності матимуть такий вигляд

$$2\frac{\partial\tilde l_{REML}(\theta)}{\partial\theta_j} = T_j + \sum_{i=1}^{M}\mathsf{trace}\left(V_i^{-1}\frac{\partial V_i}{\partial\theta_j}\right) +$$

$$+\mathsf{trace}\left((X'V^{-1}X)^{-1}X'\frac{\partial V^{-1}}{\partial\theta_j}X\right)$$

та

$$2\frac{\partial^2\tilde l_{REML}(\theta)}{\partial^2\theta_k\theta_j} = -\frac{T_kT_j}{n-p}+$$

$$+\frac{1}{\widehat{\sigma^2}}y'\frac{\partial^2V^{-1}}{\partial^2\theta_k\theta_j}y - \frac{2}{\widehat{\sigma^2}}\left(y\frac{\partial^2V^{-1}}{\partial\theta_k\theta_j}\right)'\hat\beta-$$

$$-\frac{2}{\widehat{\sigma^2}}\left(y\frac{\partial V^{-1}}{\partial\theta_k}X(X'V^{-1}X)^{-1}\right)'y\frac{\partial V^{-1}}{\partial\theta_j}X+$$

$$+\frac{2}{\widehat{\sigma^2}}\left(y\frac{\partial V^{-1}}{\partial\theta_k}X(X'V^{-1}X)^{-1}\right)'X'\frac{\partial V^{-1}}{\partial\theta_j}X\hat\beta+$$

$$+\frac{2}{\widehat{\sigma^2}}y\frac{\partial V^{-1}}{\partial\theta_j}X(X'V^{-1}X)^{-1}X'\frac{\partial V^{-1}}{\partial\theta_k}X\hat\beta-$$

$$-\frac{2}{\widehat{\sigma^2}}\left(\hat\beta X'\frac{\partial V^{-1}}{\partial\theta_j}X\right)'(X'V^{-1}X)^{-1}X'\frac{\partial V^{-1}}{\partial\theta_k}X\hat\beta+$$

$$+\frac{1}{\widehat{\sigma^2}}\hat\beta'X'\frac{\partial^2V^{-1}}{\partial^2\theta_k\theta_j}X + \sum_{i=1}^{M}\mathsf{trace}\left(\frac{\partial V^{-1}}{\partial\theta_j}\frac{\partial V}{\partial\theta_k}\right)$$

$$-\mathsf{trace}\left((X'V^{-1}X)^{-1}X'\frac{\partial V^{-1}}{\partial\theta_j}\times\right.$$

$$\left.\times X(X'V^{-1}X)^{-1}X'\frac{\partial V^{-1}}{\partial\theta_k}X\right)+$$

$$+\mathsf{trace}\left((X'V^{-1}X)^{-1}X'\frac{\partial^2V^{-1}}{\partial\theta_k\theta_j}X\right),$$

де i-й блок блочно-діагональної матриці других похідних дорівнює

$$\frac{\partial^2V_i^{-1}}{\partial^2\theta_k\theta_j} = -\frac{\partial V_i^{-1}}{\partial\theta_k}\frac{\partial V_i}{\partial\theta_j}V_i^{-1} - V_i^{-1}\frac{\partial V_i}{\partial\theta_j}\frac{\partial V_i^{-1}}{\partial\theta_k},$$

якщо $k \neq j$, та

$$\frac{\partial^2V_i^{-1}}{\partial^2\theta_k^2} = -2\frac{\partial V_i^{-1}}{\partial\theta_k}\frac{\partial V_i}{\partial\theta_k}V_i^{-1}.$$

Наступні кроки

У подальшому автори планують розширити це дослідження для багаторівневих ієрархічних лінійних моделей та інших типів кореляційних структур.

Список літератури

1. Майборода Р. Є. Регресія: лінійні моделі. ВПЦ «Київський університет», 2007. 296 с.
2. Bates D., Machler M., Bolker B., Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015. Vol. 67. Pp. 1–48.
3. Lindstrom M. J., Bates D. M. Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data.*Journal of the American Statistical Association*. 1988. Vol. 83. Pp. 1014–1022.
4. Pinheiro J. C., Bates D. M. Mixed-effect models in S and S-PLUS. Sringer : Statistics and Computing, 2000. 528 p.
5. Searle S. R., Casella G., McCulloch C. E. Variance Components. John Wiley & Sons, 2006. 536 p.
6. SAS Institute Inc. SAS/STAT® 13.1. User's Guide. Cary, NC: SAS Institute Inc, 2013.

References

1. R. Maiboroda,*Regresia: liniini modeli* (VPC "Kyivskyi universytet", 2007).
2. D. Bates, M. Machler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4б," Journal of Statistical Software. **67**, 1–48 (2015).
3. M. J. Lindstrom and D. M. Bates, "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," Journal of the American Statistical Association. **83**, 1014–1022 (1988).
4. J. C. Pinheiro and D. M. Bates, *Mixed-effect models in S and S-PLUS* (Sringer: Statistics and Computing, 2000).
5. S. R. Searle, G. Casella, and C. E. McCulloch. *Variance Components* (John Wiley & Sons, 2009).
6. SAS Institute Inc. SAS/STAT® 13.1. User's Guide (Cary, NC: SAS Institute Inc, 2013).

S. Lukashevych, R. Yamnenko

LIKELIHOOD FUNCTION DERIVATIVES FOR A LINEAR MIXED MODEL WITH COMPOUND SYMMETRY ASSUMPTION

The paper explores the properties of linear mixed models with simple random effects of the form:

$$y_i = X_i\beta + Z_i\gamma_i + \varepsilon_i, i = 1, \dots, M,$$

$$\gamma_i \sim N(0, \Psi), \varepsilon_i \sim N(0, \sigma^2 I),$$

where M is the number of distinct groups, each consisting of n_i observations. Random effects γ_i and within-group errors ε_i are independent across different groups and within the same group. β is a p-dimensional vector of fixed effects, γ_i is a q-dimensional vector of random effects, and X_i and Z_i are known design matrices of dimensions $n_i \times p$ and $n_i \times q$, of fixed and random effects respectively. Vectors ε_i represent within-group errors with a spherically Gaussian distribution.

Assuming a compound symmetry in the correlation structure of the matrix Ψ governing the dependence among within-group errors, analytical formulas for the first two partial derivatives of the profile restricted maximum likelihood function with respect to the correlation parameters of the model are derived. The analytical representation of derivatives facilitates the effective utilization of numerical algorithms like Newton-Raphson or Levenberg-Marquardt.

The restricted maximum likelihood (REML) estimation is a statistical technique employed to estimate the parameters within a mixed-effects model, particularly in the realm of linear mixed models. It serves as an extension of the maximum likelihood estimation method, aiming to furnish unbiased and efficient parameter estimates, especially in scenarios involving correlated data. Within the framework of the REML approach, the likelihood function undergoes adjustments to remove the nuisance parameters linked to fixed effects. This modification contributes to enhancing the efficiency of parameter estimation, particularly in situations where the primary focus is on estimating variance components or when the model encompasses both fixed and random effects.

Keywords: Linear mixed model, REML estimator, derivative, random effects, compound symmetry.

Матеріал надійшов 27.12.2023