РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
А В Т О М Е Т Р И Я
2004, том 40, № 1

УДК 519.23

Д. В. Лисицин
(Новосибирск)

ЕМ-АЛГОРИТМЫ ОЦЕНИВАНИЯ РЕГРЕССИОННОЙ МОДЕЛИ С МУЛЬТИПЛИКАТИВНОЙ КОВАРИАЦИОННОЙ СТРУКТУРОЙ ОШИБОК

Рассматривается регрессионная модель с коррелированными наблюдениями откликов. Группы коррелированных наблюдений имеют многомерное распределение (Стьюдента или Гаусса) с ковариационными матрицами, параметризованными в виде кронекеровского произведения нескольких положительно-определенных матриц (с точностью до скалярного сомножителя). Разрабатываются модификации ЕМ-алгоритмов [1] для оценивания параметров модели по данным с пропусками. Проводится экспериментальное исследование полученных алгоритмов.

Введение. В практике регрессионного моделирования встречаются ситуации, когда часть наблюдений откликов отсутствует. Если наблюдения в наборе данных зависимы, для их обработки требуются специальные процедуры. Один из подходов состоит в применении модификации метода максимального правдоподобия (ММП), в котором функция правдоподобия строится на основе маргинальных распределений наблюдений, реально присутствующих в наборе данных [1]. Для вычисления оценок в данном случае могут потребоваться алгоритмы, отличные от используемых при полностью комплектных наборах данных. Однако последние часто обладают положительными свойствами, которые желательно сохранить и в рассматриваемом случае. Так, в [2, 3] разработаны алгоритмы вычисления ММП-оценок параметров регрессионной модели с эллиптическим распределением ошибок и ковариационными матрицами, параметризованными в виде кронекеровского произведения нескольких положительно-определенных матриц (с точностью до скалярного сомножителя), при этом для оценивания ковариационных параметров используются достаточно простые вычислительные формулы. При построении функции правдоподобия по маргинальным распределениям ковариационная структура, основанная на кронекеровском произведении, нарушается, вследствие чего указанные вычислительные схемы не могут быть использованы.

Решить возникшую задачу можно, предварительно восстанавливая пропущенные значения и обрабатывая затем полученный полный набор данных. Различные методы восстановления пропущенных значений описаны в [1], локальные алгоритмы, относящиеся к семейству алгоритмов ZET, – в работе

[4]. К указанному подходу близки EM-алгоритмы, в которых восстанавливаются функции пропущенных данных, входящие в функцию правдоподобия полных данных. Различные модификации EM-алгоритмов используются для ММП-оценивания параметров многомерных распределений (нормального и стьюдентовского) и регрессионных моделей [1, 5–10]. В данной работе указанные алгоритмы применяются для ММП-оценивания регрессионной модели, рассмотренной в [2, 3].

Модель. Пусть проведены наблюдения за одним или несколькими откликами исследуемой системы. Пусть вектор всех наблюдений за всеми откликами можно разбить на l подвекторов (групп наблюдений) так, что наблюдения из разных групп являются независимыми, а наблюдения в одной группе коррелированы. Тогда линейно-параметризованную регрессионную модель можно представить в виде

$$Y_i = X_i\theta + E_i, \qquad i=1,\ldots,l,$$

где Y_i – вектор i-й группы измерений откликов; X_i – матрица значений регрессоров (функций входных переменных) i-й группы; θ – m-мерный вектор истинных значений неизвестных параметров; E_i – вектор ошибок наблюдений i-й группы, имеющий нулевой вектор математических ожиданий. На параметры уравнения регрессии при этом может быть наложена система линейных ограничений-равенств вида

$$R\theta = d,$$

где R – известная матрица полного строчного ранга; d – известный вектор.

Введем предположение, что вектор E_i имеет распределение Стьюдента (t-распределение) с функцией плотности

$$\frac{\nu_{\gamma(i)}^{\nu_{\gamma(i)}/2}\Gamma[(\nu_{\gamma(i)}+n_i)/2]}{\pi^{n_i/2}\Gamma(\nu_{\gamma(i)}/2)}|V_i|^{-1/2}[\nu_{\gamma(i)}+q_i]^{-(n_i+\nu_{\gamma(i)})/2},$$

где $\nu_{\gamma(i)}$ – параметр формы распределения, при этом функция $\gamma(i)$ ставит номер параметра формы в соответствие номеру группы и всего имеется n_ν различных параметров формы ν_j, $j=1,\ldots,n_\nu$; $\Gamma(z)$ – гамма-функция; n_i – размер вектора E_i; V_i – симметричная положительно-определенная псевдоковариационная матрица; $q_i = E_i^T V_i^{-1} E_i$.

Частным случаем распределения Стьюдента при $\nu_{\gamma(i)} \to \infty$ является распределение Гаусса (нормальное) с функцией плотности

$$(2\pi)^{-n_i/2}|V_i|^{-1/2}\exp[-q_i/2].$$

Параметризуем псевдоковариационные матрицы векторов E_i посредством симметричных положительно-определенных матриц U_j, $j=1,\ldots,p$, размера $k_j \geq 1$. Для псевдоковариационной матрицы i-й группы введем мультипликативную параметризацию в виде кронекеровского произведения параметров:

$$V_i = \bigotimes_{j \in J(V_i)} U_j, \qquad i=1,\ldots,l,$$

где $J(V_i)$ – упорядоченная последовательность индексов матриц U_j, посредством которых параметризована матрица V_i. Ковариационная матрица вектора E_i отличается от псевдоковариационной матрицы дополнительным скалярным множителем $d(n_i, \nu_{\gamma(i)})$ [2], поэтому она также является мультипликативно-параметризованной.

Введенная модель имеет неизвестные параметры: вектор параметров регрессии θ; матричные и скалярные параметры U_j, $j=1,\ldots,p$; параметры формы распределения ν_j, $j=1,\ldots,n_\nu$, которые требуют оценивания по данным.

ЕМ-алгоритм и его модификации. При отсутствии пропущенных наблюдений ММП-оценки доставляют максимальное значение логарифму функции правдоподобия

$$Q(\Phi) = \log f(Y \mid \Phi), \tag{1}$$

где $f(Y \mid \Phi)$ – функция плотности наблюдений; Φ – множество всех неизвестных параметров.

При наличии пропущенных наблюдений, механизм возникновения которых можно игнорировать, для получения ММП-оценок следует максимизировать функцию

$$L(\Phi) = \log f(Y_{obs} \mid \Phi), \tag{2}$$

где $f(Y_{obs} \mid \Phi)$ – функция плотности присутствующих наблюдений; Y_{obs} – совокупность значений присутствующих наблюдений [1].

Введем функцию

$$Q(\Phi \mid \tilde{\Phi}) = \mathbf{M}\{\log f(Y \mid \Phi) \mid Y_{obs}, \tilde{\Phi}\}, \tag{3}$$

где $\mathbf{M}$ – оператор взятия математического ожидания.

ЕМ-алгоритм представляет собой итерационный процесс, каждая s-я итерация которого состоит из Е- и М-шагов. На Е-шаге определяется функция $Q(\Phi \mid \Phi^{(s-1)})$ аргумента Φ. На М-шаге находится значение $\Phi^{(s)}$ как решение задачи $\max\limits_{\Phi} Q(\Phi \mid \Phi^{(s-1)})$. Итерации ЕМ-алгоритма не уменьшают значение функции $L(\Phi)$, и ЕМ-алгоритм сходится в общем случае к локальному максимуму или точке перегиба функции $L(\Phi)$. Поэтому рекомендуется выполнять ЕМ-алгоритм несколько раз с разными начальными значениями параметров. В случае если решение на М-шаге невозможно получить в явном виде (требуется итерационный процесс), используется обобщенный ЕМ-алгоритм – GEM-алгоритм (generalized EM), в котором М-шаг состоит в поиске значения $\Phi^{(s)}$, удовлетворяющего условию

$$Q(\Phi^{(s)} \mid \Phi^{(s-1)}) \geq Q(\Phi^{(s-1)} \mid \Phi^{(s-1)}).$$

Когда М-шаг ЕМ- или GEM-алгоритма является сложным для выполнения, его можно заменить последовательностью шагов условной максимизации – последовательностью CM-шагов (conditional/constrained maximization), каждый из которых максимизирует функцию $Q(\Phi \mid \Phi^{(s-1)})$ по Φ при зафиксированных значениях некоторых функций от Φ. Последовательность CM-шагов при этом должна обеспечивать максимизацию во всем параметри-

ческом пространстве. Данная модификация EM-алгоритма называется ECM-алгоритмом и предложена в [5]. В простейшем случае множество параметров разбивается на несколько подмножеств и максимизация на некотором CM-шаге производится по одному из подмножеств при фиксированных значениях остальных параметров. Подобную стратегию удобно применять, когда оптимизация по отдельным подмножествам параметров более проста, чем оптимизация по всем параметрам, однако в общем случае она не всегда приводит к успеху. Указанный подход аналогичен методу покомпонентного оценивания, использованному в [2, 3] для вычисления ММП-оценок параметров рассматриваемой модели по комплектному набору данных.

Модификация EM-алгоритма, предложенная в [6], называется ECME-алгоритмом (ECM either). В этом случае некоторые (или все) CM-шаги ECM-алгоритма, которые называются здесь CMQ-шагами, заменяются CM-шагами условной максимизации исходной функции $L(\Phi)$ (CML-шагами), причем перед последовательностью CMQ-шагов должен быть выполнен E-шаг. Данная модификация может иметь существенно более быструю сходимость итерационного процесса в сравнении с EM- и ECM-алгоритмами.

Для ускорения сходимости предназначен и PX-EM-алгоритм (parameter expansion) [7]. Здесь модель $f(Y \mid \Phi)$ заменяется расширенной моделью $f_*(Y \mid \Psi)$ с параметром $\Psi = (\Phi_*, \gamma)$, где Φ_* играет ту же роль в $f_*(Y \mid \Psi)$, что Φ в $f(Y \mid \Phi)$, а γ – дополнительный параметр, значение которого в исходной модели фиксировано и равно γ_0. При этом должны выполняться два условия: во-первых, не должна изменяться модель присутствующих данных в том смысле, что для всех Ψ существует функция редукции F такая, что плотность условного распределения $Y_{obs} \mid \Psi$ есть $f\{Y_{obs} \mid \Phi = F(\Psi)\}$; во-вторых, не должна изменяться при $\gamma = \gamma_0$ модель полных данных в том смысле, что для всех Φ выполняется $f_*\{Y \mid \Psi = (\Phi_*, \gamma_0)\} = f\{Y \mid \Phi = \Phi_*\}$. Эти условия определяют, что из $\Phi_1 \neq \Phi_2$ следует $\Psi_1 \neq \Psi_2$ и что для всех Φ существует по крайней мере одно значение Ψ, при котором плотность условного распределения $Y_{obs} \mid \Psi$ равна $f\{Y_{obs} \mid \Phi = F(\Psi)\}$. Пусть $\Psi^{(0)} = (\Phi^{(0)}, \gamma_0)$ – начальное значение, тогда s-я итерация PX-EM-алгоритма состоит из следующих шагов: на PX-E-шаге определяется функция

$$Q_*(\Psi \mid \Phi^{(s-1)}) = \mathbf{M}\{\log f_*(Y \mid \Psi) \mid Y_{obs}, \Phi_* = \Phi^{(s-1)}, \gamma = \gamma_0\},$$

на PX-M-шаге находится $\Psi^{(s)}$ как решение задачи $\max\limits_{\Psi} Q_*(\Psi \mid \Phi^{(s-1)})$, затем производится редукция $\Phi^{(s)} = F[\Psi^{(s)}]$.

Оценивание параметров. EM-алгоритмы оценивания параметров при t-распределении ошибок наблюдений строятся исходя из N/I-представления данного распределения [8–10]. Условное распределение $Y_i \mid \tau_i$ является нормальным: $Y_i \mid \tau_i \sim N\left(X_i\theta, \frac{1}{\tau_i} V_i\right)$, ненаблюдаемые величины τ_i имеют гамма-распределение с плотностью $G\left(\frac{\nu_{\gamma(i)}}{2}, \frac{\nu_{\gamma(i)}}{2}\right)$, где $G(\alpha, \beta) = \beta^{\alpha} \tau_i^{\alpha-1} \times$ $\times \exp[-\beta\tau_i] \Gamma^{-1}(\alpha)$. В PX-EM-модификации при некоторых частных видах параметризации псевдоковариационной матрицы используется расширенная модель, в которой величины τ_i имеют распределение с плотностью

$G\left(\frac{\nu_{\gamma(i)}}{2}, \frac{\nu_{\gamma(i)}}{2\gamma}\right)$, где γ – дополнительный скалярный параметр [7, 10]. При мультипликативной ковариационной структуре общего вида функция редукции не может быть построена, однако по аналогии с данным алгоритмом скалярные ковариационные параметры могут быть представлены и оценены как параметры второго аргумента функции плотности τ_i.

Пусть множество ковариационных параметров разделено на два подмножества, причем во второе подмножество входят только скалярные параметры. Обозначим множества индексов ковариационных параметров первого и второго подмножеств как J_1 и J_2 соответственно. Будем использовать N/I-представление распределения Стьюдента вида $Y_i \mid \tau_i \sim N\left(X_i\theta, \frac{1}{\tau_i}\Omega_i\right)$ при τ_i, имеющей гамма-распределение с плотностью $G\left(\frac{\nu_{\gamma(i)}}{2}, \frac{\nu_{\gamma(i)}\lambda_i}{2}\right)$, где $\Omega_i = \frac{1}{\lambda_i} V_i$, а λ_i – произведение скалярных ковариационных параметров, отнесенных ко второму подмножеству.

Если часть групп наблюдений имеет нормальное распределение, то последнее представление невозможно, для таких групп будем использовать модель $Y_i \sim N(X_i\theta, \lambda_i\Omega_i)$.

Введем обозначения: $Y_{i,\,obs}$, $Y_{i,\,mis}$ – векторы присутствующих и отсутствующих наблюдений i-й группы (заметим, что значения $Y_{i,\,mis}$ являются неизвестными); $X_{i,\,obs}$, $X_{i,\,mis}$ – подматрицы матрицы X_i, соответствующие присутствующим и отсутствующим наблюдениям i-й группы; $E_{i,\,obs}$ – подвектор вектора E_i, соответствующий присутствующим данным; $n_{i,\,obs}$ – размер вектора $Y_{i,\,obs}$. Далее, обозначим подматрицы матрицы V_i, соответствующие векторам $Y_{i,\,obs}$, $Y_{i,\,mis}$, как $V_{i,\,obs} = V_{i,\,obs,\,obs}$, $V_{i,\,obs,\,mis}$, $V_{i,\,mis,\,obs}$, $V_{i,\,mis} = V_{i,\,mis,\,mis}$. Аналогичные обозначения введем для подматриц матрицы Ω_i, т. е. справедливо

$$\operatorname{cov}\begin{bmatrix} Y_{i,\,obs} \\ Y_{i,\,mis} \end{bmatrix} = d(n_i, \nu_{\gamma(i)}) \begin{bmatrix} V_{i,\,obs} & V_{i,\,obs,\,mis} \\ V_{i,\,mis,\,obs} & V_{i,\,mis} \end{bmatrix} =$$

$$= d(n_i, \nu_{\gamma(i)})\lambda_i \begin{bmatrix} \Omega_{i,\,obs} & \Omega_{i,\,obs,\,mis} \\ \Omega_{i,\,mis,\,obs} & \Omega_{i,\,mis} \end{bmatrix}.$$

Поскольку значения τ_i ненаблюдаемы, их следует обрабатывать как пропущенные.

Запишем функции (1)–(3) с точностью до константных слагаемых, при этом для (1), (3) воспользуемся представлением распределения Стьюдента в виде N/I-распределения:

$$L(\Phi) = \sum_{j=1}^{n_\nu} \sum_{i \in J(\nu_j)} \left\{ \log\Gamma\left(\frac{n_{i,\,obs} + \nu_j}{2}\right) + \frac{\nu_j}{2}\log\nu_j - \log\Gamma\left(\frac{\nu_j}{2}\right) - \right.$$

$$\left. - \frac{1}{2}\log|V_{i,\,obs}| - \frac{n_{i,\,obs} + \nu_j}{2}\log(\nu_j + q_{i,\,obs}) \right\} - \sum_{i \in J_N} \left\{ \frac{1}{2}\log|V_{i,\,obs}| + \frac{1}{2} q_{i,\,obs} \right\},$$

$$Q(\Phi)=\sum_{i\in J_t}\left[\frac{n_i}{2}\log\tau_i-\frac{1}{2}\log|\Omega_i|-\frac{\tau_i}{2}(Y_i-X_i\theta)^T\Omega_i^{-1}(Y_i-X_i\theta)\right]-$$

$$-\sum_{i\in J_N}\left[\frac{1}{2}\log|\lambda_i\Omega_i|+\frac{q_i}{2}\right]+$$

$$+\sum_{j=1}^{n_\nu}\sum_{i\in J(\nu_j)}\left[\frac{\nu_j}{2}\log\frac{\nu_j\lambda_i}{2}+\left(\frac{\nu_j}{2}-1\right)\log\tau_i-\frac{\nu_j\lambda_i\tau_i}{2}-\log\Gamma\left(\frac{\nu_j}{2}\right)\right],$$

$$Q(\Phi\,|\,\tilde{\Phi})=\sum_{i\in J_t}\left[\frac{n_i\tilde{u}_i}{2}-\frac{1}{2}\log|\Omega_i|-\frac{1}{2}\operatorname{tr}\Omega_i^{-1}(\tilde{w}_i\tilde{S}_i+\tilde{C}_i^{\Omega})\right]-$$

$$-\sum_{i\in J_N}\left[\frac{1}{2}\log|\lambda_i\Omega_i|+\frac{1}{2}\operatorname{tr}V_i^{-1}(\tilde{S}_i+\tilde{C}_i^{V})\right]+$$

$$+\sum_{j=1}^{n_\nu}\sum_{i\in J(\nu_j)}\left[\frac{\nu_j}{2}\log\frac{\nu_j\lambda_i}{2}+\left(\frac{\nu_j}{2}-1\right)\tilde{u}_i-\frac{\nu_j\lambda_i\tilde{w}_i}{2}-\log\Gamma\left(\frac{\nu_j}{2}\right)\right],$$

где $J(\nu_j)$ – множество номеров групп, для распределения которых ν_j является параметром формы; $q_{i,\,obs}=E_{i,\,obs}^T V_{i,\,obs}^{-1}E_{i,\,obs}$; J_N, J_t – множества индексов групп, имеющих распределения Гаусса и Стьюдента соответственно; $\tilde{u}_i=$ $=\mathbf{M}\{\log\tau_i\,|\,Y_{obs},\tilde{\Phi}\}$; $\tilde{w}_i=\mathbf{M}\{\tau_i\,|\,Y_{obs},\tilde{\Phi}\}$; $\tilde{S}_i=(\tilde{Y}_i-X_i\theta)(\tilde{Y}_i-X_i\theta)^T$; $\tilde{Y}_i=$ $=\mathbf{M}\{Y_i\,|\,Y_{obs},\tilde{\Phi}\}$: $\tilde{Y}_{i,\,obs}=Y_{i,\,obs}$, $\tilde{Y}_{i,\,mis}=X_{i,\,mis}\tilde{\theta}+\tilde{\Omega}_{i,\,mis,\,obs}\tilde{\Omega}_{i,\,obs}^{-1}(Y_{i,\,obs}-X_{i,\,obs}\tilde{\theta})$; $\tilde{C}_i^{V}$ – матрица размера $n_i\times n_i$, (j,k)-й элемент которой равен соответствующему элементу матрицы $\tilde{V}_{i,\,mis}-\tilde{V}_{i,\,mis,\,obs}\tilde{V}_{i,\,obs}^{-1}\tilde{V}_{i,\,obs,\,mis}$, если в векторе Y_i отсутствуют j- и k-й элементы одновременно, и нулю в остальных случаях; $\tilde{C}_i^{\Omega}$ – матрица, аналогичная $\tilde{C}_i^{V}$, но построенная по матрице $\tilde{\Omega}_i$.

Можно показать, что условное распределение $\tau_i\,|\,Y_{obs},\tilde{\Phi}$ есть гамма-распределение с плотностью $G\left(\frac{\tilde{\nu}_{\gamma(i)}+n_{i,\,obs}}{2},\,\tilde{\lambda}_i\frac{\tilde{\nu}_{\gamma(i)}+\tilde{q}_{i,\,obs}}{2}\right)$, где $\tilde{q}_{i,\,obs}=(Y_{i,\,obs}-$ $-X_{i,\,obs}\tilde{\theta})^T\tilde{V}_{i,\,obs}^{-1}(Y_{i,\,obs}-X_{i,\,obs}\tilde{\theta})$, тогда имеем

$$\tilde{w}_i=\frac{1}{\tilde{\lambda}_i}\frac{\tilde{\nu}_{\gamma(i)}+n_{i,\,obs}}{\tilde{\nu}_{\gamma(i)}+\tilde{q}_{i,\,obs}},\qquad \tilde{u}_i=\psi\left(\frac{\tilde{\nu}_{\gamma(i)}+n_{i,\,obs}}{2}\right)-\log\frac{\tilde{\lambda}_i(\tilde{\nu}_{\gamma(i)}+\tilde{q}_{i,\,obs})}{2},$$

где $\psi(z)$ – дигамма-функция.

Поиск условного минимума будем производить методом неопределенных множителей Лагранжа, т. е. минимизируя функцию

$$Q_R=Q+\varphi^T(R\theta-d),$$

где φ – вектор неизвестных множителей Лагранжа.

Систему уравнений правдоподобия после ряда преобразований можно представить в виде

$$\hat{\theta} = \hat{\theta}_0 + H^{-1} R^T (R H^{-1} R^T)^{-1} (d - R\hat{\theta}_0),$$

$$(\hat{U}_j)_{\alpha\beta} = \frac{k_j}{N_j} \left[\sum_{i \in J(U_j) \cap J_t} \mathrm{tr} \hat{\Omega}^{-1}_{i,[j,\alpha,\beta]} (\tilde{w}_i \tilde{S}_i + \tilde{C}_i^{\Omega}) + \right.$$

$$\left. + \sum_{i \in J(U_j) \cap J_N} \mathrm{tr} \hat{V}^{-1}_{i,[j,\alpha,\beta]} (\tilde{S}_i + \tilde{C}_i^{V}) \right], \quad j \in J_1, \quad \alpha, \beta = 1, \ldots, k_j,$$

$$\hat{U}_j = \left[\sum_{i \in J(U_j) \cap J_t} \hat{\nu}_{\gamma(i)} - \sum_{i \in J(U_j) \cap J_N} n_i \right] \times$$

$$\times \left[\sum_{i \in J(U_j) \cap J_t} \hat{\nu}_{\gamma(i)} \tilde{w}_i \prod_{\substack{k \in J(\lambda_i) \\ k \neq j}} \hat{U}_k - \frac{1}{\hat{U}_j^2} \sum_{i \in J(U_j) \cap J_N} \prod_{\substack{k \in J(\lambda_i) \\ k \neq j}} \frac{1}{\hat{U}_k} \mathrm{tr} \hat{\Omega}_i^{-1} (\tilde{S}_i + \tilde{C}_i^{V}) \right]^{-1}, \quad j \in J_2, \tag{4}$$

$$\log \frac{\hat{\nu}_j}{2} - \psi\left(\frac{\hat{\nu}_j}{2}\right) + \Delta_j = 0, \quad j = 1, \ldots, n_\nu,$$

где

$$H = \sum_{i \in J_t} \tilde{w}_i X_i^T \hat{\Omega}_i^{-1} X_i + \sum_{i \in J_N} X_i^T \hat{V}_i^{-1} X_i;$$

$$\hat{\theta}_0 = H^{-1} \left[\sum_{i \in J_t} \tilde{w}_i X_i^T \hat{\Omega}_i^{-1} \tilde{Y}_i + \sum_{i \in J_N} X_i^T \hat{V}_i^{-1} \tilde{Y}_i \right];$$

$J(U_j)$ – множество индексов групп наблюдений, в число параметров псевдоковариационных матриц которых входит матрица U_j; $N_j = \sum_{i \in J(U_j)} n_i$; $\hat{\Omega}^{-1}_{i,[j,\alpha,\beta]}$, $\hat{V}^{-1}_{i,[j,\alpha,\beta]}$ – матрицы, совпадающие с $\hat{\Omega}_i^{-1}$, $\hat{V}_i^{-1}$ соответственно, в которых вместо сомножителя $\hat{U}_j$ стоит матрица того же размера, все элементы которой, за исключением (α,β)-го, равны нулю, а (α,β)-й элемент равен единице; $J(\lambda_j)$ – множество индексов ковариационных параметров, являющихся сомножителями в λ_j;

$$\Delta_j = 1 + \frac{1}{\mu(\nu_j)} \sum_{i \in J(\nu_j)} \{\tilde{u}_i - \hat{\lambda}_i \tilde{w}_i + \log \hat{\lambda}_i\};$$

$\mu(\nu_j)$ – мощность множества $J(\nu_j)$.

Для поиска значений оценок неизвестных параметров воспользуемся ECM-алгоритмом. Разделим множество неизвестных параметров на следующие подмножества: $\{\theta_i,\ i=1,2,\ldots,m\}, \{[U_1]_{i,j=1}^{k_1}\}, \{[U_2]_{i,j=1}^{k_2}\}, \ldots, \{[U_p]_{i,j=1}^{k_p}\}$, $\{\nu_1\}, \{\nu_2\}, \ldots, \{\nu_{n_\nu}\}$, тогда М-шаг алгоритма состоит из ряда CMQ-шагов, на каждом из которых осуществляется оптимизация по параметрам одного из подмножеств при фиксированных значениях параметров остальных подмножеств. В этом случае итерационными являются шаги вычисления оценок параметров формы, а при наличии нормально распределенных групп – и параметров U_j, $j \in J_2$.

В работах [6, 10] рассматривается ECME-алгоритм для оценивания параметров распределения Стьюдента, в котором CMQ-шаг оценивания параметров формы заменяется CML-шагом. Решаемое на нем уравнение имеет вид

$$\sum_{i \in J(\nu_j)} \left\{ \psi\left(\frac{n_{i,obs} + \hat{\nu}_j}{2} \right) - \log \frac{\hat{\nu}_j + \hat{q}_{i,obs}}{2} - \frac{n_{i,obs} + \hat{\nu}_j}{\hat{\nu}_j + \hat{q}_{i,obs}} \right\} +$$

$$+ \mu(\nu_j) \left\{ 1 + \log \frac{\hat{\nu}_j}{2} - \psi\left(\frac{\hat{\nu}_j}{2} \right) \right\} = 0,$$

где $\hat{q}_{i,obs} = (Y_{i,obs} - X_{i,obs}\hat{\theta})^T \hat{V}_{i,obs}^{-1} (Y_{i,obs} - X_{i,obs}\hat{\theta})$. Для поиска корня данного уравнения методом Ньютона используется производная по $\hat{\nu}_j$ левой части уравнения, которая имеет вид

$$\sum_{i \in J(\nu_j)} \left\{ \frac{1}{2} \psi'\left(\frac{n_{i,obs} + \hat{\nu}_j}{2} \right) - \frac{1}{\hat{\nu}_j + n_{i,obs}} \right\} + \mu(\nu_j) \left\{ \frac{1}{\hat{\nu}_j} - \frac{1}{2} \psi'\left(\frac{\hat{\nu}_j}{2} \right) \right\} +$$

$$+ \sum_{i \in J(\nu_j)} \frac{(n_{i,obs} - \hat{q}_{i,obs})^2}{(\hat{\nu}_j + n_{i,obs})(\hat{\nu}_j + \hat{q}_{i,obs})^2}.$$

Если данное выражение не является отрицательным, при использовании метода Ньютона последнюю сумму в [10] рекомендуется отбрасывать.

Вычисление по формуле (4) при наличии групп с нормально распределенными ошибками может приводить к отрицательным значениям на некоторых итерациях. В результате целесообразно второе подмножество формировать из ковариационных параметров стьюдентовских групп, возможно также присутствие небольшого числа нормальных групп. В последнем случае при наличии на очередной итерации отрицательного значения необходимо модифицировать алгоритм оценивания. Так, можно использовать CML-шаг, приводящий к итеративному процессу решения уравнения

$$\hat{U}_j = \sum_{i \in J(U_j)} \frac{\hat{\nu}_{\gamma(i)} + n_{i,obs}}{\hat{\nu}_{\gamma(i)} + \hat{q}_{i,obs}} \prod_{\substack{k \in J(\lambda_i) \\ k \neq j}} \frac{1}{\hat{U}_k} \hat{E}_{i,obs}^T \hat{\Omega}_{i,obs}^{-1} \hat{E}_{i,obs} \Bigg/ \sum_{i \in J(U_j)} n_{i,obs},$$

где $\hat{E}_{i,\,obs} = Y_{i,\,obs} - X_{i,\,obs}\hat{\theta}$, например, на основе метода последовательных приближений.

Рассмотренные способы построения алгоритмов для оценивания параметров могут быть применены и при отсутствии пропущенных наблюдений откликов. В этом случае пропусками будут только значения латентной переменной τ. В отличие от описанного подхода непосредственная максимизация функции $L(\Phi)$ с использованием покомпонентного оценивания приводит к алгоритму, имеющему внутренние итеративные процессы на всех шагах одной внешней итерации.

Экспериментальное исследование. С целью проверки работоспособности предложенных алгоритмов оценивания параметров было проведено экспериментальное исследование методом статистических испытаний, основанным на многократном генерировании выборки с последующим нахождением по ней оценок параметров.

Данные генерировались в соответствии с моделью

$$y = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3,$$

где $\theta_1 = 10$, $\theta_2 = -5$, $\theta_3 = 2$, $\theta_4 = 0{,}2$, $x_i \in [-1,1]$.

При оценивании на параметры было наложено ограничение $\theta_2 + \theta_3 = -3$, а значение оценки параметра формы предполагалось лежащим в интервале 0,01–1000.

Показателями, отражавшими полезность разработанных алгоритмов, являлись: t – время в секундах, затрачиваемое на поиск решения, и κ – число итераций.

Выборка состояла из 64 групп по 8 наблюдений в каждой. Использовались мультипликативные структуры псевдоковариационной матрицы $V_i = U_1 U_2 \rho_i$, $i = 1,\ldots,64$, $k_1 = 4$, $k_2 = 2$, где

$$U_1 = \sqrt{0{,}0557125}\begin{bmatrix} 10 & \sqrt{2{,}5} & \sqrt{2{,}5} & \sqrt{2{,}5} \\ \sqrt{2{,}5} & 1{,}0 & 0{,}5 & 0{,}5 \\ \sqrt{2{,}5} & 0{,}5 & 1{,}0 & 0{,}5 \\ \sqrt{2{,}5} & 0{,}5 & 0{,}5 & 1{,}0 \end{bmatrix};$$

$$U_2 = \sqrt{0{,}0557125}\begin{bmatrix} 10 & \sqrt{2{,}5} \\ \sqrt{2{,}5} & 1{,}0 \end{bmatrix};$$

при следующих мультипликативных параметризациях скаляра ρ_i:

1) $\rho_i = \delta_1$, $i = 1,\ldots,64$,

2) $\rho_i = \delta_{1+[(i-1)/16]}$, $i = 1,\ldots,64$, $[\bullet]$ – целая часть числа,

3) $\rho_i = \delta_{1+[(i-1)/16]}\sigma_{1+[(i-1)/16]}$, $i = 1,\ldots,64$,

4) $\rho_i = \delta_{1+[(i-1)/8]}$, $i = 1,\ldots,64$,

5) $\rho_i = \delta_{1+[(i-1)/8]}\sigma_{1+[(i-1)/8]}$, $i = 1,\ldots,64$,

6) $\rho_i = \delta_{1+[(i-1)/16]}\sigma_{1+[(i-1)/8]}$, $i = 1,\ldots,64$,

7) $\rho_i = \delta_{1+[(i-1)/4]}\sigma_{1+[(i-1)/8]}, \rho_{i+32} = \delta_{1+[(i-1)/4]}\sigma_{5+[(i-1)/8]}, i=1,\ldots,32,$

8) $\rho_i = \delta_{1+[(i-1)/4]}\sigma_{1+[(i-1)/8]}, \rho_{i+16} = \delta_{1+[(i-1)/4]}\sigma_{3+[(i-1)/8]},$
$\rho_{i+32} = \delta_{1+[(i-1)/4]}\sigma_{5+[(i-1)/8]}, \rho_{i+48} = \delta_{1+[(i-1)/4]}\sigma_{7+[(i-1)/8]}, i=1,\ldots,16,$

9) $\rho_i = \delta_1\sigma_{1+[(i-1)/8]}, i=1,\ldots,64$

(все скалярные параметры δ_j и σ_j имели единичные значения).

Пропуски в наборе данных генерировались случайно, при этом вероятность появления пропуска в каждом наблюдении равнялась 0,1.

Сравнивались четыре алгоритма. В первых двух алгоритмах – ECM и ECME – все ковариационные параметры принадлежали первому подмножеству, в последних двух алгоритмах – ECM-G и ECME-G – скалярные ковариационные параметры δ_j принадлежали второму подмножеству.

В качестве начальных значений в алгоритмах использовались оценка по методу наименьших квадратов для θ и единичные матрицы для ковариационных параметров. Критерием останова алгоритма было относительное изменение функции (2) на соседних итерациях, меньшее 10^{-7}.

Рассмотрим результаты исследования случая, когда ошибки имели распределение Стьюдента с параметром формы $\nu = 8$, а начальное значение параметра формы было равно 1. В табл. 1 приведены средние значения показателей t и к, полученные по пятидесяти реализациям выборки, и выделены наименьшие значения показателей.

Результаты, приведенные в таблице, показывают существенное преимущество алгоритма ECME перед ECM. Преимущество алгоритма ECME-G проявлялось в тех случаях, когда каждая группа из кластера групп, имеющих

Таблица 1

Показатели полезности алгоритмов при $\nu = 8$

Номер ковариационной структуры	ECM-алгоритм		ECME-алгоритм		ECM-G-алгоритм		ECME-G-алгоритм	
	t	к	t	к	t	к	t	к
1	0,5442	31,72	0,3308	18,8	0,5498	33,08	**0,1826**	**10,6**
2	0,7824	48,36	**0,3074**	**18,5**	0,7686	49,44	0,3106	19,6
3	1,0476	48,36	0,4102	18,5	1,0204	48,92	**0,2314**	**10,8**
4	1,8106	106,84	**0,3204**	**18,28**	1,7482	107,64	0,6596	40,14
5	2,2388	106,84	0,3928	18,28	2,1394	106,34	**0,2316**	**11,14**
6	2,2256	106,84	0,3874	18,28	2,1442	107,24	**0,286**	**13,96**
7	3,2064	153,02	**0,3754**	**17,54**	3,0786	152,96	1,4542	71,9
8	2,9444	141,18	**0,3802**	**17,88**	2,8278	141,76	1,2162	60,46
9	2,2898	106,84	0,4008	18,28	2,23	107,76	**0,3126**	**14,7**

Таблица 2

Показатели полезности алгоритмов при $\nu = 1$

Номер ковариационной структуры	ECM-алгоритм		ECME-алгоритм		ECM-G-алгоритм		ECME-G-алгоритм	
	t	к	*t*	к	*t*	к	*t*	к
1	1,1528	66,66	1,2254	68,18	0,214	12,74	**0,1932**	**10,82**
4	1,2336	69,52	1,2796	69,7	0,2512	14,44	**0,22**	**12,1**
5	1,506	69,52	1,5496	69,7	0,3034	14,4	**0,2612**	**11,86**
6	1,5136	69,52	1,564	69,7	1,1284	53,36	**1,1272**	**52,12**
7	1,5	69,04	1,5422	68,94	**1,1498**	54,46	1,1596	**53,58**
9	1,5386	69,52	1,588	69,7	**1,2538**	58,38	1,2584	**57,08**

определенный ковариационный параметр из второго подмножества, обязательно имела ковариационный параметр из первого подмножества, присутствующий только у групп данного кластера, причем не обязательно у всех. Сравнение характеристик ECME-алгоритма для ковариационной структуры 4 и ECME-G-алгоритма для ковариационных структур 5, 6, 9 позволяет сделать вывод, что увеличение количества параметров, оцениваемых по гамма-распределению, уменьшает время работы алгоритма.

Результаты исследования при ошибках, имеющих распределение Стьюдента с параметром формы $\nu = 1$ и начальным значением параметра формы 1000, отражены в табл. 2, аналогичной табл. 1. В этом случае алгоритм ECME несколько уступает алгоритму ECM, однако G-модификации всегда имеют преимущество перед стандартными алгоритмами, причем в ряде случаев это преимущество существенно.

Заключение. Полученные в работе модификации EM-алгоритмов предназначены для оценивания параметров регрессионной модели с ошибками, имеющими мультипликативную ковариационную структуру и распределения Стьюдента и Гаусса, в условиях как наличия, так и отсутствия пропусков. Проведенное исследование позволяет рекомендовать для использования ECME- и ECME-G-алгоритмы, в которых параметры, оцениваемые по гамма-распределению, дублируют параметры, оцениваемые по нормальному распределению, причем последние являются наиболее гибкими, т. е. каждый из них присутствует у минимального количества групп.

СПИСОК ЛИТЕРАТУРЫ

1. **Литтл Р. Дж. А., Рубин Д. Б.** Статистический анализ данных с пропусками. М.: Финансы и статистика, 1990.
2. **Лисицин Д. В.** Регрессионная модель с эллиптическим распределением и мультипликативной ковариационной структурой ошибок // Науч. вест. НГТУ. 2001. № 2(11). С. 67.

3. **Денисов В. И., Лисицин Д. В.** Оценивание параметров регрессионной модели с эллиптическим распределением и мультипликативной ковариационной структурой ошибок // Сиб. журн. индустр. матем. 2002. V, № 3(11). С. 92.

4. **Загоруйко Н. Г.** Прикладные методы анализа данных и знаний. Новосибирск: Изд-во ИМ СО РАН, 1999.

5. **Meng X.-L., Rubin D. B.** Maximum likelihood estimation via the ECM algorithm: A general framework // Biometrika. 1993. **80**. P. 267.

6. **Liu C., Rubin D. B.** The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence // Biometrika. 1994. **81**. P. 633.

7. **Liu C., Rubin D. B., Wu Y. N.** Parameter expansion to accelerate EM: the PX-EM algorithm // Biometrika. 1998. **85**. P. 755.

8. **Lange K. L., Little R. J. A., Taylor J. M. G.** Robust statistical modeling using the t-distribution // Journ. Amer. Statist. Assoc. 1989. **84**. P. 881.

9. **Pinheiro J. C., Liu C., Wu Y. N.** Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t-distribution // Journ. Comput. and Graph. Statist. 2001. **10**. P. 249.

10. **Liu C.** ML estimation of the multivariate t-distribution and the EM algorithms // Journ. Multivar. Anal. 1997. **63**. P. 296.

Новосибирский государственный технический университет

Поступила в редакцию 25 апреля 2003 г.