Оценки максимального правдоподобия в условиях цензурирования на основе EM-алгоритма

Коробейников Антон Иванович

28 мая 2008 г.

Содержание

1. Введение

1.1. Оценки максимального правдоподобия

Оценивание посредством максимального правдоподобия и статистические процедуры, основанные на правдоподобии, представляют огромный интерес для задач статистики и анализа данных. Метод максимального правдоподобия является универсальным методом, обладающим широким спектром полезных статистических свойств. Стоит отметить, что метод применим не только при традиционном выборочном подходе, но и при использовании байесовского подхода к анализу данных. Байесовские оценки обычно эквивалентны оценкам максимального правдоподобия с ограничениями. Метод максимального правдоподобия повсеместно используется во всех областях, где применяется аппарат статистики.

Всюду далее будем предполагать, что наблюдаемая случайная величина Y имеет плотность $g(y;\theta)$, где $\theta \in \Theta$ — неизвестный вектор параметров. Основная цель метода — максимизация функции правдоподобия $L(\theta) = g(Y;\theta)$ как функции аргумента θ при фиксированном значении наблюдаемой случайной величины Y. Это, в свою очередь, равносильно нахождению корней уравнения

$$\frac{\partial L(\theta)}{\partial \theta} = 0, \tag{1.1}$$

или, заменяя функцию правдоподобия её логарифмом,

$$\frac{\partial \log L(\theta)}{\partial \theta} = 0. \tag{1.2}$$

Задачей оценивания методом максимального правдоподобия является нахождение оценки $\hat{\theta}$ таким образом, что полученная последовательность корней уравнения (1.2) являлась бы состоятельной и асимптотически эффективной. Известно, что при наложении определенных условий регулярности на плотность g, такая последовательность корней существует [2]. С вероятностью, стремящейся к единице, эти корни соответствуют точкам локального максимума функции правдоподобия во внутренности пространства параметров Θ.

Как правило, функция правдоподобия для стандартных статистических моделей имеет глобальный максимум во внутренности Θ. Таким образом, последовательность корней уравнения (1.2) с требуемыми асимптотическими свойствами естественным образом получается посредством выбора корня $\hat{\theta}$ доставляющего максимум $L(\theta)$ глобально во всей области Θ. Конечно, в данном случае, $\hat{\theta}$ является оценкой максимального правдоподобия.

Отметим, что далее оценкой максимального правдоподобия (ОМП) $\hat{\theta}$ будет называться корень уравнения (1.2) даже если он не доставляет функции правдоподобия $L(\theta)$ глобального максимума. Известно, что для некоторых статистических моделей, включающих в себя смеси распределений [7], функция правдоподобия может оказаться неограниченной. Однако, для подобных моделей при выполнении обычных условий регулярности, может существовать последовательность корней уравнения (1.2) со всеми требуемыми свойствами состоятельности, эффективности и асимптотической нормальности [5].

Если функция правдоподобия (или ее логарифм) имеет профиль (по параметрам) близкий к квадратичному, как в случае независимых случайных величин, имеющих нормальное распределение, то ее максимум может быть легко найден посредством решения системы линейных уравнений относительно неизвестных параметров. Однако, на практике часто оказывается, что профиль функции правдоподобия весьма далек от квадратичного. Это, в свою очередь, приводит в методе максимального правдоподобия ко всевозможным эффектам, связанным с нелинейностью. Примерами подобных ситуация являются:

- Оценивание функции от неизвестных параметров.
- Несмотря на возможную линейную структуру профиль функция правдоподобия далек от квадратичного из-за, например, ошибок измерений, не имеющих нормального распределения, зависимости между наблюдениями, пропущенных наблюдений.

Традиционным способом нахождения оценок максимального правдоподобия в нелинейном случае являются всевозможные модификации алгоритма Ньютона-Рафсона. При некоторых допущениях относительно функции $L(\theta)$ и достаточно аккуратно выбранном начальном приближении, последовательность приближений к корню $\{\hat{\theta}^{(n)}\}$, полученная из метода Ньютона-Рафсона, демонстрирует локальную квадратичную сходимость к решению θ^* уравнения (1.2). Квадратичную скорость сходимости обычно выделяют среди основных достоинств этого метода.

Однако, в реальных приложениях это семейство методов оптимизации оказывается весьма неудобным как в аналитическом, так и в вычислительном плане даже в простейших случаях [6, 9]. EM-алгоритм является разумной альтернативой этим методам во многих важных приложениях и широко используется для итеративного вычисления оценок максимального правдоподобия. Подобного рода задачи естественным образом возникают при анализе данных с пропусками или при наличии неполной информации о выборке.

1.2. EM-алгоритм и данные с пропусками

При применении различных статистических методов для обработки реальных данных достаточно часто можно столкнуться с ситуацией, когда оценивание неизвестных параметров распределения методом максимального правдоподобия затруднено из-за сложной структуры функции правдоподобия, приводящей к труднорешаемым задачам оптимизации. При этом сложности могут быть как аналитического, так и вычислительного характера. Типичным примером подобных задач является обработка цензурированных, сгруппированных или усеченных данных, многомерных данных с пропусками, данных, полученных из смеси распределений.

Во многих подобных случаях часто возможно переформулировать статистическую задачу в терминах тех же параметров, но с «дополнительно введенными» данным так, что полученная задача оценивания методом максимального правдоподобия становится вполне решаемой как с вычислительной, так и с аналитической точки зрения. Введенные данные можно назвать *полными данными*, а доступную выборку — *неполными данными*; функции правдоподобия — *полной* и *неполной* соответственно.

EM-алогоритм решает (сложную) задачу вычисления оценок максимального правдоподобия для неполных данных посредством формулирования соответствующей задачи для полных данных и использования простоты вычисления ОМП в этом случае для вычисления оценок параметров исходной задачи. Введенные данные могут рассматриваться как *пропущенные* по отношению к реальной задаче с неполными данными. Отметим, что пропущенные данные не обязательно могут быть пропущенными в нормальном смысле этого слова, а являться, в некотором роде, удобным техническим инструментом для решения задачи. В связи с этим, термин *данные с пропусками* достаточно широко используется для обозначения всевозможных статистических моделей, включающих смеси распределений, свертку, группировку, цензурирование, усечение и, собственно, непосредственно пропущенные наблюдения.

EM-алгоритм является итеративным алгоритмом, каждая итерация которого состоит из двух шагов: шага математического ожидания (expectation step, E-шаг) и шага максимизации (maximization step, M-шаг). Краткую историю EM-алгоритма можно найти в работе [6]. Собственно название «EM-алгоритм» было предложено в статье [3], где были собраны воедино и обобщены ранние формулировки алгоритма, применяемые для решения конкретных реальных задач, и была представлена общая формулировка этого алгоритма для нахождения ОМП параметров достаточно широкого спектра статистических задач. С тех пор EM-алгоритм применялся для решения как всевозможных общих статистических задач, включая разделение смесей распределений, оценка компонент дисперсий, факторный анализ, так и специальных, в таких областях как генетика, нейронные сети, медицинская обработка изображений и пр.

2. EM-алгоритм

Всюду далее, будем обозначать через X (неизвестный) вектор полных данных, а через Z — вектор «пропущенных» данных. В отдельных случаях исследуемая задача на первый взгляд может не относиться к классу задач с неполными данными, тем не менее, процедура вычисления оценок максимального правдподобия может быть существенно упрощена за счет искусственного переформулирования задачи и применения процедур построения оценок в условиях неполных данных.

Основной предпосылкой для подобного рода переформулирования служит тот факт, что для многих классических статистических задач в случае отсутствия неполных данных функция правдоподобия имеет достаточно простой вид. В свою очередь, EM-алгоритм существенно использует факт простоты построения оценок максимального правдоподобия для наблюдений без пропусков.

2.1. Шаги алгоритма

Обозначим через $g_c(x;\,\theta)$ плотность случайной величины X. Тогда логарифм функции правдоподобия для X (если бы она в действительности наблюдалась) имеет вид

$$\log L_c(\theta) = \log g_c(X;\,\theta). \tag{2.1}$$

EM-алгоритм решает задачу решения уравнения правдоподобия для неполных данных (1.2) неявно, используя для этого логарифм функции правдоподобия для полных данных. Так как случайная величина X не наблюдается, а, значит, не наблюдается и $\log L_c(\theta)$, то используется условное математическое ожидание $\log L_c(\theta)$ по отношению к наблюдаемой случайной величине Y. Приведем $(k+1)$-ю итерацию EM-алгоритма:

E-шаг: Вычислить функцию $Q\left(\theta;\,\theta^{(k)}\right)$, где

$$Q\left(\theta;\,\theta^{(k)}\right) = \mathbf{E}_{\theta^{(k)}}\left(\log L_c(\theta)\,|\,Y\right). \tag{2.2}$$

M-шаг: Выбрать в качестве следующего $\theta^{(k+1)}$ точку максимума функции $Q\left(\theta;\,\theta^{(k)}\right)$:

$$Q\left(\theta^{(k+1)};\,\theta^{(k)}\right) \geqslant Q\left(\theta;\,\theta^{(k)}\right) \quad \forall \theta \in \Theta. \tag{2.3}$$

Всюду далее символом $\mathbf{E}_{\theta^{(k)}}\left(\cdot\right)$ будет обозначаться математическое ожидание с использованием вектора параметров $\theta^{(k)}$.

E- и M-шаги алгоритма последовательно выполняются до тех пор, пока последовательность $\theta^{(k)}$ не достигнет стационарной точки. Этот момент может быть определен при помощи подходящего критерия остановки, например, если $\left\|\theta^{(k+1)} - \theta^{(k)}\right\| \leqslant \varepsilon$ для некоторого $\varepsilon > 0$ и выбранной нормы $\|\cdot\|$, либо $\left|L\left(\theta^{(k+1)}\right) - L\left(\theta^{(k)}\right)\right| < \varepsilon$.

2.2. Случай экспоненциального семейства

Оба шага итерации EM-алгоритма принимают особенно простой вид, когда плотность $g_c(x;\,\theta)$ принадлежит экспоненциальному семейству:

$$g_c(x;\,\theta) = b(x)\exp\left[c(\theta)^T t(x)\right]/a(\theta), \tag{2.4}$$

где $t(x)$ — вектор достаточных статистик размера $k \times 1$ $(k > d)$, $c(\theta)$ — вектор-функция размера $k \times 1$ от d-мерного вектора параметров θ, $b(x)$ и $a(\theta)$ — скалярные функции. Экспоненциальному семейству принадлежат основные классические распределения, такие как многомерное нормальное, пуассоновское, мультиномиальное и другие.

Для плотности, принадлежащей экспоненциальному семейству, E-шаг алгоритма может быть записан в виде:

$$Q\left(\theta;\,\theta^{(k)}\right) = \mathbf{E}_{\theta^{(k)}}\left(\log b(X)\,|\,Y\right) + c(\theta)^T t^{(k)} - \log a(\theta), \tag{2.5}$$

где $t^{(k)} = \mathbf{E}_{\theta^{(k)}}\left(t(X)\,|\,Y\right)$ — оценка достаточной статистики. Во время M-шага необходимо максимизировать функцию $Q\left(\theta;\,\theta^{(k)}\right)$, но $\mathbf{E}_{\theta^{(k)}}\left(\log b(X)\,|\,Y\right)$ от θ не зависит, поэтому итерация алгоритма упрощается до:

E-шаг: Вычислить

$$t^{(k)} = \mathbf{E}_{\theta^{(k)}}\left(t(X)\,|\,Y\right). \tag{2.6}$$

M-шаг: Вычислить

$$\theta^{(k+1)} = \underset{\theta\in\Theta}{\arg\max}\left[c(\theta)^T t^{(k)} - \log a(\theta)\right]. \tag{2.7}$$

2.3. Обобщенный EM-алгоритм

Часто максимум функции $Q\left(\theta;\,\theta^{(k)}\right)$, находимый во время M-шага, можно получить в явной форме. Однако, бывают такие ситуации, когда даже численное нахождения этого максимума сопряжено с большими трудностями. Для решения этой проблемы в работе [3] был предложен обобщенный EM-алгоритм (Generalized EM, GEM), в котором M-шаг модифицирован таким образом, что требуется найти лишь такое значение $\theta^{(k+1)}$, что:

$$Q\left(\theta^{(k+1)};\,\theta^{(k)}\right) \geqslant Q\left(\theta^{(k)};\,\theta^{(k)}\right). \tag{2.8}$$

Таким образом, требование глобальной максимизации заменяется гораздо более мягким требованием нахождения хотя бы одного значения $\theta^{(k+1)}$, увеличивающим значение функции $Q\left(\theta;\,\theta^{(k)}\right)$ по отношению к значению в точке $\theta = \theta^{(k)}$.

Заметим, кроме того, что итерация EM-алогритма (GEM-алогритма) задает неявным образом отображение $\theta \to M(\theta)$ пространства параметров Θ в себя:

$$\theta^{(k+1)} = M(\theta^{(k)}),\, k = 0, 1, \ldots. \tag{2.9}$$

Отображение M называется *EM-отображением*. Оно будет использоваться при обсуждении вопросов сходимости алгоритма.

2.4. Сходимость EM-алгоритма

Обозначим через $k(x\,|\,y;\,\theta) = g_c(x;\,\theta)/g(y;\,\theta)$ условную плотность X при $Y = y$. Тогда логарифм функции правдоподобия для полных данных может быть записан в виде

$$\log L_c(\theta) = \log g_c(x;\,\theta) = \log L(\theta) + \log k(x\,|\,y;\,\theta). \tag{2.10}$$

Применяя оператор условного математического ожидания $\mathbf{E}_{\theta^{(k)}}\left(\cdot\,|\,Y\right)$ к обеим частям уравнения (2.10) и обозначая $H(\theta;\,\theta^{(k)}) = \mathbf{E}_{\theta^{(k)}}\left(\log k(X\,|\,Y;\,\theta)\,|\,Y\right)$, имеем

$$Q\left(\theta;\,\theta^{(k)}\right) = \log L(\theta) + H(\theta;\,\theta^{(k)}). \tag{2.11}$$

Из формулы (2.11) непосредственно выводим, что

$$\log L(\theta^{(k+1)}) - \log L(\theta^{(k)}) = \left[Q\left(\theta^{(k+1)};\,\theta^{(k)}\right) - Q\left(\theta^{(k)};\,\theta^{(k)}\right)\right] - \left[H(\theta^{(k+1)};\,\theta^{(k)}) - H(\theta^{(k)};\,\theta^{(k)})\right]. \tag{2.12}$$

Применяя неравенство Йенсена, имеем $H(\theta^{(k+1)};\,\theta^{(k)}) \leqslant H(\theta^{(k)};\,\theta^{(k)})$. А из (2.3) или (2.8) следует, что $Q\left(\theta^{(k+1)};\,\theta^{(k)}\right) - Q\left(\theta^{(k)};\,\theta^{(k)}\right) \geqslant 0$.

Таким образом, значение функции правдоподобия неубывает после итерации EM- или GEM-алгоритма:

$$L(\theta^{(k+1)}) \geqslant L(\theta^{(k)}), \quad k = 0, 1, 2, \ldots. \tag{2.13}$$

Следствием (2.13) является так называемое свойство *самосостоятельности* EM-алгоритма. Так, если множество значений функции правдоподобия $\left\{L(\theta^{(k)})\right\}$ ограничено, то $L(\theta^{(k)})$ монотонно стремится к некоторому значению L^*. Здесь естественным образом встает вопрос о том, какие условия требуются для того, чтобы L^* было стационарным значением и когда это значение является если не глобальным максимумом, то хотя бы локальным. Известны примеры, когда EM-алгоритм сходится к *глобальному минимуму* или же к седловой точке [6]. Также естественным является вопрос с сходимости последовательности $\left\{\theta^{(k)}\right\}$ к ОМП $\hat{\theta}$ (то есть к состоятельному корню уравнения правдоподобия).

До представления EM-алогритма в общем виде в статье [3] в литературе встречались результаты касающиеся сходимости частных случаев EM-алгоритма. Так, например, в работе [1] были представлены результаты о сходимости EM-алгоритма, примененного к оценке параметров скрытых марковских моделей. В самой работе [3] были приведены некоторые результаты, относящиеся к сходимости алгоритма в общем случае. В наибольшей общности детальное изучение свойств сходимости EM-алгоритма было проведено в работе [11]. В ней было показано, что в случае, когда полные данные принадлежат экспоненциальному семейству с компактным пространством параметров Θ и когда функция Q удовлетворяет некоторым не слишком сильным условиям дифференцируемости, последовательность приближений $\left\{\theta^{(k)}\right\}$ сходится к стационарной точке функции правдоподобия (не обязательно к максимуму). Если функция $L(\theta)$ имеет несколько стационарных точек, то сходимость к стационарной точке определенного вида (точке локального или глобального максимума, седловой точке) зависит от начального приближения $\theta^{(0)}$. Если $L(\theta)$ является унимодальной на Θ и удовлетворяет тому же условию дифференцируемости, то любая последовательность $\left\{\theta^{(k)}\right\}$ сойдется к ОМП $\hat{\theta}$ не зависимо от начального приближения.

В частности, одним из свойств сходимости EM-алгоритма является

$$\log L\left(M(\theta)\right) \geqslant \log L(\theta), \tag{2.14}$$

где равенство достигается тогда и только тогда, когда

$$Q\left(M(\theta);\, \theta\right) = Q\left(\theta;\, \theta\right), \quad k\left(X\,|\,Y;\, M(\theta)\right) = k\left(X\,|\,Y;\, \theta\right). \tag{2.15}$$

Тем самым, функция правдоподобия увеличивается на каждой итерации алгоритма, до тех пор, пока не будет достигнута неподвижная точка. Если $\hat{\theta}$ является ОМП, то есть $L(\hat{\theta}) \geqslant L(\theta), \forall\theta \in \Theta$, то, очевидно, $\log L(M(\hat{\theta})) = \log L(\hat{\theta})$. Таким образом, оценки максимального правдоподобия необходимо являются неподвижными точками EM-отображения.

Если функция правдоподобия ограничена (а так получается во многих важных случаях), то последовательность $\left\{\theta^{(k)}\right\}$ порождает неубывающую ограниченную последовательность $\left\{\log L(\theta^{(k)})\right\}$, которая, очевидно, сходится при $k \to \infty$.

Ясно, что неподвижные точки EM-отображения, вообще говоря, не всегда являются ОМП. В работе [11] были получены достаточно широкие достаточные условия того, что EM-алгоритм сходится к ОМП.

Так, например, предположим, что (см. [4]):

1. $$\left[\frac{\partial Q\left(\theta;\, \theta^{(k)}\right)}{\partial \theta}\right]_{\theta=\theta^{(k+1)}} = 0,$$
2. Последовательность $\left\{\theta^{(k)}\right\}$ сходится к некоторому значению θ^* и функция $\log k(x \,|\, y;\, \theta)$ обладает достаточной гладкостью.

Тогда $[\partial \log L(\theta)/\partial\theta]_{\theta=\theta^*} = 0$. Мы видим, что нет никакой гарантии того, что последовательность $\left\{\theta^{(k)}\right\}$ сойдется к точке глобального максимума функции правдоподобия. Если функция правдоподобия является многоэкстремальной, то EM-алгоритм, сходится к точке локального максимума, вообще говоря, зависящей от начального приближения $\theta^{(0)}$.

В работе [10] условия сходимости были расширены на пространства параметров с ограничениями. Кроме того, там же были выведены более жесткие условия, гарантирующие сходимость EM-алгоритма к ОМП.

2.5. Скорость сходимости EM-алгоритма

Скорость сходимости алгоритма имеет важное значение для его применения на практике. Как правило, скорость сходимости EM-алгоритма меньше квадратичнной, присущей алгоритмам типа Ньютона-Рафсона. В работе [3] было показано, что скорость сходимости EM-алгоритма линейна и, кроме того, зависит от «объема информации», сохраненной в наблюдаемых данных. То есть, сравнивая с задачей построения оценок по данным без пропусков, сходимость может быть достаточно медленной.

Рассмотрим введенное ранее EM-отображение. Если последовательность $\theta^{(k)}$ сходится к некоторой величине θ^* и $M(\theta)$ непрерывно, то θ^* является неподвижной точкой отображения M, то есть $\theta^* = M(\theta^*)$. Тогда, раскладывая $\theta^{(k+1)} = M(\theta^{(k)})$ по формуле Тейлора вблизи точки θ^*, имеем:

$$\theta^{(k+1)} = M(\theta^{(k)}) \approx M(\theta^*) + J(\theta^*)\left(\theta^{(k)} - \theta^*\right) = \theta^* + J(\theta^*)\left(\theta^{(k)} - \theta^*\right), \qquad (2.16)$$

откуда

$$\theta^{(k+1)} - \theta^* \approx J(\theta^*) \left(\theta^{(k)} - \theta^*\right). \tag{2.17}$$

Здесь $J(\theta)$ — якобиан отображения $M(\theta)$. Таким образом, в окрестности точки θ^* EM-алгоритм является фактически линейным отображением с матрицей $J(\theta^*)$, которая во многих случаях отлична от нуля. Значение якобиана $J(\theta^*)$ обычно называют матричной скоростью сходимости. Для самого вектора параметра θ можно ввести меру скорости сходимости, определяемую как

$$r = \lim_{k\to\infty} \frac{\left\|\theta^{(k+1)} - \theta^*\right\|}{\left\|\theta^{(k)} - \theta^*\right\|}, \tag{2.18}$$

где $\|\cdot\|$ — любая норма в d-мерном Евклидовом пространстве $\mathbb{R}^d$. В работе [9] было показано, что при выполнении некоторых условий регулярности величина r совпадает с наибольшим по модулю собственным числом матрицы $J(\theta^*)$. Так как большее значение r означает более медленную сходимость, то непосредственно скорость сходимости обычно определяют как $s = 1 - r$ (смотри, например, [8]).

3. Пример применения EM-алгоритма

Приведем несложный пример, иллюстрирующих применение EM-алгоритма для решения практических задач.

При анализе данных типа времени жизни наблюдаемой величной является случайное время до наступления какого-либо события («наработки на отказ»). Во многих случаях это событие не происходит за все время наблюдений, либо индивид выходит из-под наблюдения до окончания эксперимента. Такие данные называются «цензурированными»: доступна лишь информация о том, что интересуемое время наступления события больше некоторой известной величины.

Таким образом, наблюдается случайный вектор $Y = ((c_1, \delta_1), (c_2, \delta_2), \ldots, (c_n, \delta_n))$, где $\delta_j = 0$ или 1 в зависимости от того, является j-е наблюдение T_j цензурированным в точке c_j, или нет. Так, если T_j не цензурировано, то $c_j = t_j$, если же $t_j > c_j$, то наблюдение цензурировано в точке c_j.

Будем предполагать для T простую экспоненциальную модель с плотностью

$$f(t;\, \mu) = \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right),\, t > 0,\, \mu > 0. \tag{3.1}$$

Обозначим через s количество нецензурированных наблюдений. Перенумеруем выборку так, что нецензурированные наблюдения получили номера меньше, чем цензурированные. Тогда логарифм функции правдоподобия для μ принимает вид

$$\log L(\mu) = -s \log \mu - \sum_{j=1}^{n} \frac{c_j}{\mu}. \tag{3.2}$$

Приравнивая производную $\log L(\mu)$ по μ в (3.2) к нулю, получаем выражение для ОМП $\hat{\mu}$:

$$\hat{\mu} = \frac{1}{s} \sum_{j=1}^{n} c_j. \tag{3.3}$$

Ясно, что нет никакой необходимости в каком-либо итеративном вычислении ОМП. Тем не менее, продемонстрируем, как EM-алгоритм может быть применен в данном случае (и как он упрощается за счет того, что плотность (3.1) принадлежит к регулярному экспоненциальному семейству).

Вектор полных данных X можно представить в виде $X = (t_1, \ldots, t_s, Z)$, где вектор $Z = (t_{s+1}, \ldots, t_n)$ содержит $n - s$ ненаблюдаемых времен отказа цензурированной части выборки. Функция правдоподобия для полных данных имеет вид:

$$\log L_c(\mu) = -n \log \mu - \frac{1}{\mu} \sum_{j=1}^{n} t_j. \tag{3.4}$$

Так как $\log L_c(\mu)$ является линейной функцией от ненаблюдаемой части выборки Z, то Е-шаг алгоритма сводится к простой замене Z на ее текущее условное математическое ожидание по отношению к наблюдаемой части выборки Y. Используя свойство «отсутствия памяти» экспоненциального распределения, выводим, что условное распределение случайной величины $T_j - c_j$ при условии $T_j > c_j$ является экспоненциальным со средним μ. Таким образом, имеем:

$$\mathbf{E}_{\mu^{(k)}}\left(T_j \,|\, Y\right) = \mathbf{E}_{\mu^{(k)}}\left(T_j \,|\, T_j > c_j\right) = c_j + \mu^{(k)}, \quad j = s+1, \ldots, n. \tag{3.5}$$

Соответственно, функция Q принимает вид

$$Q\left(\mu;\, \mu^{(k)}\right) = -n\log\mu - \frac{1}{\mu}\left(\sum_{j=1}^{n} c_j + (n-s)\mu^{(k)}\right). \tag{3.6}$$

После М-шага, имеем

$$\mu^{(k+1)} = \frac{1}{n}\left(\sum_{j=1}^{n} c_j + (n-s)\mu^{(k)}\right). \tag{3.7}$$

Положим в (3.7) $\mu^{(k+1)} = \mu^{(k)} = \mu^*$. Решая полученное уравнение относительно μ^*, получаем, что $\mu^* = \hat{\mu}$. Таким образом, последовательность $\left\{\mu^{(k)}\right\}$ при $k \to \infty$ имеет единственно возможной предельной точкой ОМП $\hat{\mu}$.

Из выражения для логарифма функции правдоподобия для полных данных (3.4) видно, что T принадлежит экспоненциальному семейству вида (2.4) с каноническим параметром μ^{-1} и достаточной статистикой $t(X) = \sum_{j=1}^{n} T_j$. Из (3.5) видно, что Е-шаг алгоритма для экспоненциального семейства заключается в вычислении $t^{(k)} = \sum_{j=1}^{n} c_j + (n-s)\mu^{(k)}$. М-шаг заключается в нахождении $\mu^{(k+1)}$ как значения μ, удовлетворяющего уравнению

$$t^{(k)} = \mathbf{E}_{\mu}\left(t(X)\right) = n\mu.$$

Несложно видеть, что решение этого уравнения в точности задается формулой (3.7).

4. Заключение

EM-алгоритм обладает некоторыми полезными свойствами, такими как:

1. Численная устойчивость. Каждая итерация алгоритма увеличивает значение функции правдоподобия.

2. При достаточно общих условиях алгоритм сходится.

3. Алгоритм прост в реализации как аналитически, так и численно. Как правило, программная реализация алгоритма проста и требует мало памяти. Если функция правдоподобия вычисляется несложным образом, то можно следить за монотонным увеличением значения функции правдоподобия на каждой итерации. Это позволяет легко контролировать сходимость и обнаруживать возможные ошибки в программной реализации алгоритма.

4. Как правило, сложность одной итерации невелика, что, впрочем, компенсируется необходимость выполнить достаточно большое количество итераций (по сравнению с другими алгоритмами).

5. Алгоритм может использоваться для оценивания пропущенных данных.

Естественно, алгоритм обладает и рядом недостатков:

1. Сам по себе алгоритм не предоставляет оценки ковариационной матрицы оценок параметров. Однако, этот недостаток может быть достаточно просто преодолен за счет некоторой модификации алгоритма [6].

2. Иногда скорость сходимости алгоритма невелика.

3. В некоторых случаях может не существовать замкнутой аналитической формы для отдельных шагов алгоритма.

5. Список литературы

[1] Baum, L.E, Petrie, T., Soules, G., Weiss, N. *A maximisation technique occurring in the statistical analysis of probabalistic functions of Markov processes* // Ann. of Math. Stat., **1970**, *41*, P. 164-171.

[2] Cramér, H. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey, *1946*. [Есть перевод на русском языке: Крамер Г.Н. *Математические методы статистики*. М.:Мир, *1975* г.]

[3] Dempster, A.P., Laird, N.M., Rubin, D.B. *Maximum likelihood data from incomplete data via the EM algorithm* // J. R. Statis. Soc. Ser. B., **1977**, *39*, P. 1-38.

[4] Little, R.J.A., Rubin, D.B. *Statistical Analysis with Missing Data*, Wiley, New York, *2002*.

[5] McLachlan, G.J., Basford, K.E. *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker, New York, *1988*.

[6] McLachlan, G.J., Krishnan, T. *The EM Algorithm and Extensions*. Wiley, New York, *1997*.

[7] McLachlan, G.J., Peel, D. *Finite Mixture Models*. Wiley, New York, *2000*.

[8] Meng, X.L. *On the rate of convergence of the ECM algorithm* // Ann. of Stat., **1994**, *22*, P. 326-339.

[9] Meng, X.L., van Dyk, D. *The EM algorithm – an old song sung to a fast new tune* // J. R. Statis. Soc. Ser. B., **1997**, *59*, P. 511-567.

[10] Nettleton, D. *Convergence properties of the EM algorithm in constrained parameter spaces* // Canad. J. of Stat., **1999**, *27*, P. 639-648.

[11] Wu, C.F.J. *On the convergence properties of the EM algorithm* // Ann. of Stat., **1983**, *11*, P. 95-103.