

*Московский государственный университет имени М.В. Ломоносова*

*Факультет вычислительной математики и кибернетики*

*Кафедра математической статистики*

**ЕМ алгоритм**

**Курсовая работа**

Студента 4 курса

*Е.Г. Птицына*

**Научный руководитель:**

д.ф.-м.н., профессор

*В.Ю. Королёв*

# Оглавление

[1. Оглавление 2](#_Toc7006577)

[2. Введение 3](#_Toc7006578)

[2.1. Метод максимального правдоподобия 3](#_Toc7006579)

[3. Описание ЕМ алгоритма 5](#_Toc7006580)

[3.1. Е- и М- шаги 5](#_Toc7006581)

[3.2. Монотонность ЕМ-алгоритма 6](#_Toc7006582)

[3.3. ЕМ алгоритм как проксимальный 6](#_Toc7006583)

[3.4. Скорость сходимости 8](#_Toc7006584)

[3.5. Выбор начального приближения 9](#_Toc7006585)

[3.6. Правила остановки 10](#_Toc7006586)

[4. Свойства ЕМ Алгоритма 11](#_Toc7006587)

[5. Применение ЕМ алгоритма 11](#_Toc7006588)

[5.1. Разделение конечных смесей нормальных распределений 11](#_Toc7006589)

[6. Модификации ЕМ алгоритма 13](#_Toc7006590)

[6.1. Медианные модификации ЕМ-алгоритма 13](#_Toc7006591)

[6.1.1. Первая медианная модификация 14](#_Toc7006592)

[6.1.2. Вторая медианная модификация 14](#_Toc7006593)

[6.2. Stochastic EM(SEM) алгоритм 15](#_Toc7006594)

[6.3. Классификационный ЕМ-алгоритм (Classification EM, CEM) 16](#_Toc7006595)

[6.4. Обобщённый ЕМ-алгоритм (Generalized EM, GEM) 16](#_Toc7006596)

[7. Заключение 16](#_Toc7006597)

[8. Список Литературы 17](#_Toc7006598)

# Введение

Данная работа посвящена изучению ЕМ-алгоритма и его применения в прикладной статистике. Цель курсовой работы дать теоретическое описание работы ЕМ алгоритма, привести примеры его использования.

EM алгоритмом принято называть довольно работоспособную схему построения процедур итерационного типа для численного решения задачи поиска экстремума целевой функции в разнообразных задачах оптимизации. В частности, в прикладной статистике эта схема достаточно эффективна при поиске оценок максимального правдоподобия и родственных им ситуациях, когда функция правдоподобия имеет сложную структуру, из-за которой другие методы оказываются неэффективными или вообще неприменимы, например в задачах с неполными данными.

Он базирован на идее решения набора подзадач, которые получаются путем изменения оригинальных наблюдаемых переменных (неполные данные) с набором дополнительных переменных, которые не наблюдаемы или недоступны для пользователя, эти дополнительные данные называют *скрытыми данными*. ЕМ-алгоритм тесно связан с специальным подходом к оценке отсутствующих данных. Последние обновляются по их прогнозируемым значениям используя начальные оценки параметров. Затем параметры переоцениваются, и так далее итерационно до сходимости. На каждой итерации ЕМ-алгоритма, есть два шага называемые E-шаг (Expectation шаг) и М-шаг (Maximization шаг). Название «ЕМ-алгоритм» было дано Dempster et al (1977) в их фундаментальной публикации.

ЕМ-алгоритм имеет хорошие свойства, такие как численная устойчивость, надежная сходимость к локальным максимумам на всем пространстве и простота исполнения. Однако, ЕМ-алгоритм имеет ограничения. В его базовом виде ЕМ-алгоритме отсутствует встроенная процедура для вычисления ковариационной матрицы оценки параметров. Так же в некоторых случаях EM алгоритм очень медленно сходится. Более того, определенный комплекс задач с неполными данными приводит к неразрешимым E- и M-шагам.

## Метод максимального правдоподобия

Оценки максимального правдоподобия и основанные на функции правдоподобия выводы являются важнейшими в статистической теории и анализе данных. Метод максимального правдоподобия является базовым методом с хорошими свойствами. Это наиболее часто используемая методика в частотном анализе, и она может быть в равной степени применима к поиску апостериорного распределения в Байесовском анализе. (C. P. Robert, Bayesian Computational Methods)

Зачастую Байесовские решения оправдывают с помощью функции правдоподобия и оценок максимального правдоподобия (MLE), и Байесовские решения похожи на оценки правдоподобия со штрафом. Оценка максимального правдоподобия является универсальным методом и широко используется в любой области, где используются статистические методы.

Предположим, что наблюдаемые данные имеет плотность , где вектор, содержащий неизвестные параметры в постулированной форме для плотности **.** Наша цель максимизировать функцию правдоподобия как функцию , над параметрическим пространством . Так, что,

Или эквивалентно для логарифма правдоподобия,

Цель ОМП - оценить так, чтобы определить последовательность корней (2.1) согласованную и асимптотически эффективную. Подобная последовательность существует при подходящих регулярных условиях (Cramér 1946). Почти наверно, эти корни соответствуют локальному максимуму на . Для оценочных моделей в общем, функция правдоподобия имеет глобальный максимум на . Типичная последовательность корней, (2.1) с желаемыми асимптотическими свойствами, получается взяв как корень, который глобально максимизирует ; в этом случае, является Оценкой Максимального Правдоподобия. Далее будем называть – ОМП даже в ситуациях, когда она может не давать глобальный максимум правдоподобия. Несомненно, в некоторых примерах на смесях (McLachlan and Peel 2000, Chap. 3), функция правдоподобия не ограниченна. Однако, для этих моделей, при обычных условиях регулярности, могут существовать последовательности корней (2.1) со свойствами согласованности, эффективности и асимптотический нормальности. (McLachlan and Basford 1988, Chap. 12).

Когда функция правдоподобия или логарифмическая функция правдоподобия являются квадратичными в параметрах, как в случае независимых нормально распределенных наблюдений, ее максимум может быть получен решением системы линейных уравнений с параметрами. Однако, часто на практике функция правдоподобия не является квадратичной, что приводит к нелинейным проблемам в оценке максимального правдоподобия. Например: (а) модели приводят к математическим ожиданиям нелинейным в параметрах; (б) несмотря на возможную линейную структуру, функция правдоподобия не квадратичная по параметрам по причине не «нормальных» ошибок, отсутствующих данных или зависимостях.

Традиционно ОМП в этих ситуациях производится численно итерационным методом, решая уравнения, например методом Ньютона-Рафсона (NR) и его вариантами, такими как метод Фишера. При разумных предположениях на и достаточно точных начальные значениях, последовательность полученная методом Ньютона-Рафсона имеет локальную квадратичную сходимость к решению для (2.1). Квадратичная сходимость рассматривается как основное преимущество метода Ньютона-Рафсона. Но в применении эти методы могут быть утомительны аналитически и численно даже в сравнительно простых случаях. Смотри McLachlan and Krishnan (2008, Sect. 1.3) и Meng and van Dyk (1997). ЕМ-алгоритм предлагает лучшую альтернативу по многим критериям. Сейчас это популярный инструмент для итерационного нахождения ОМП в различных проблемах с отсутствующими или неполными данными.

# Описание ЕМ алгоритма

## Е- и М- шаги

В результате действия ЕМ-алгоритма, представляющего собой итерационную процедуру, вычисляется последовательность значений параметра . Если задано некоторое значение , то вычисление следующего значения можно условно подразделить на два этапа, аббревиатура наименования которых и дала название всей процедуре. Опишем эти этапы.

**E-шаг.** Вычисление математического ожидания (Expectations). Вычисляется значение вектора скрытых переменных по текущему приближению вектора.

Определим функцию как условное математическое ожидание логарифма полной функции правдоподобия при известном значении наблюдаемой компоненты **X*:***

В этом определении *θ* являетсяаргументом функции , **X** и – параметры, так что соотношении (3.1) символ означает усреднение по **Y** относительно меры .

При известном значении **X = *x*** функцию можно вычислить по формуле

**М-шаг**. Производится максимизация функции правдоподобия и находится следующее приближение вектора θ.

Итерационный процесс останавливается в соответствии с заранее согласованным критерием остановки. Например, заранее выбирается какая-нибудь метрика и фиксируется малое положительное число ε. Процесс останавливается на m-ом шаге, если .

Заметим, что иногда название «ЕМ-алгоритм», данное в бумаге Dempster et al. (1977), объясняют не упомянутой выше аббревиатурой английских слов *Expectation-Maximization,* но возводят к термину *Estimation-Maximization.* (например (Айвазян и др., 1989)). По всей вероятности, первый термин все же лучше отражает суть ЕМ-алгоритма.

## Монотонность ЕМ-алгоритма

Свойство монотонности ЕМ-алгоритма было впервые установлено в работе (Шлезингер, 1965). Впоследствии это свойство обобщенных и модифицированных версий ЕМ-алгоритма систематически исследовалось в работах (Dempster, Laird and Rubin, 1977), (Everitt and Hand, 1981), (Boyles,1983), (Wu, 1983), (Redner and Walker, 1984), (Jordan and Xu, 1996), (Xuand Jordan, 1996).

Недавно в работах (Neal and Hinton, 1998) и (Chretien and Hero, 2000) было замечено, что ЕМ-алгоритм принадлежит к классу так называемых проксимальных алгоритмов (или PP-алгоритмов, от английского термина Proximal Point algorithms) (в статье (Neal and Hinton, 1998) отмечено соответствующее ключевое свойство ЕМ-алгоритма, но при этом ЕМ-алгоритм формально не идентифицирован как PP-алгоритм). Это замечание существенно упрощает исследование свойства монотонности ЕМ-алгоритма. При этом главную роль играет следующее представление функции .

Из соотношения вытекает, что

По этому

Согласно принятой выше терминологии, обычная статистическая процедура поиска оценок максимального правдоподобия направлена на максимизацию по логарифма неполной функции правдоподобия, который при известном значении равен первому слагаемому в правой части (3.2.1)

## ЕМ алгоритм как проксимальный

Рассмотрим расстояние Кульбака–Лейблера (Kullback and Leibler,1951) (см. также, например, (Кульбак, 1967), (Cover and Thomas, 1991)) между условными плотностями и , определяемое как условное математическое ожидание логарифма отношения правдоподобия при известном значении наблюдаемой компоненты относительно меры :

Расстояние Кульбака–Лейблера можно вычислить по формуле

При этом, применяя неравенство Иенсена, в силу вогнутости логарифмической функции легко убедиться, что

Таким образом, расстояние Кульбака–Лейблера удовлетворяет условиям:

1.   
2.

Положив

мы таким образом замечаем, что свойства 1 и 2 расстояния Кульбака–Лейблера соответствуют условиям, которые определяют штрафную функцию в определении проксимального алгоритма.

Осталось убедиться, что с и расстоянием Кульбака–Лейблера в качестве штрафной функции соотношение (5.3.11), определяющее проксимальный алгоритм, трансформируется в соотношение, определяющее М-этап ЕМ-алгоритма. Действительно, при таких и имеем

Заметим, что второе слагаемое в правой части последнего соотношения (равное дифференциальной энтропии условного распределения ) не зависит от . Поэтому

Но выражение в фигурных скобках в правой части оказывается в точности равным функции , откуда вытекает требуемое соотношение

Другими словами, рекуррентные соотношения  
и

задают одну и ту же последовательность, то есть ЕМ-алгоритм является специальным проксимальным алгоритмом. А так как любой проксимальный алгоритм обладает свойством монотонности, то свойство монотонности оказывается присущим и ЕМ-алгоритму в том смысле, что, если последовательность вычисляется в соответствии с правилами, определяющими ЕМ-алгоритм, то  
то есть

[1]

## Скорость сходимости

Скорость сходимости ЕМ-алгоритма обычно меньше, чем квадратичная, доступная с Ньютоновскими методами. В работе (Dempster et al., 1977) показано, что скорость сходимости ЕМ алгоритма линейна и скорость зависит от доли информации в наблюдаемых данных. Следовательно, в сравнении с задачей с полными данными, если большая часть данных отсутствует, сходимость может быть достаточно медленной.

Определим отображение из параметрического пространства ***Ω*** в само себя, таким, что . Функция ***М*** называется ЕМ отображением.

Если сходится к какой-то точке и непрерывна, тогда стационарная точка алгоритма; такая что, должна удовлетворять условиям . Разложив в ряд Тейлора в точке , в окрестности получим что

Где есть матрица Якоби для , имеющая (i,j)-й элемент равный

Где и *d* размерность **.** Таким образом, в окрестности , ЕМ алгоритм представляет собой, по существу, линейную итерацию с матрицей отношений поскольку обычно не равен нулю. По этой причине часто называют матрицей сходимости. Для вектора , мера фактически наблюдаемой скорости сходимости есть глобальная скорость сходимости определенная как

где это любая норма на *d*-мерном Евклидовом пространстве . Отмечено, что наблюдаемая скорость сходимости равна наибольшему собственному значению при определённых условиях регулярности (Meng and van Dyk 1997). Поскольку большое значение r подразумевает медленную сходимость, глобальная скорость сходимости определена как (Meng, 1994); см. так же McLachlan and Krishnan (2008, Sect. 3.9).

## Выбор начального приближения

ЕМ алгоритм будет сходиться медленно если выбрано плохое начальное значение . Действительно в некоторых случаях, когда функция правдоподобия стремится в бесконечность, на краю параметрического пространства, последовательность оценок генерированная ЕМ алгоритмом может расходиться если выбран слишком близко к границе. Так же с приложением, где уравнение правдоподобия имеет несколько корней, соответствующих локальным максимумам, ЕМ алгоритм следует применять из широкого набора начальных значений в поиске всех локальных максимумов. Вариант ЕМ алгоритма (Wright and Kennedy, 2000) использует метод интервального анализа, для отыскания стационарных точек логарифма функции правдоподобия внутри любого обозначенного региона параметрического пространства; см. также McLachlan and Krishnan (2008, Sect. 7.9).

Различные способы определения начального значения были рассмотрены специально в рамках модели смесей. С EMMIX программой (McLachlan and Peel 2000, pp. 343–344), значение начального параметра может быть получено, автоматически используя случайные части данных, или алгоритм кластеризации k-средних, или иерархический метод кластеризации. С случайными начальными значениями, эффект центральной предельной теоремы делает параметры изначально похожими, по крайней мере в больших выборках. С программой EMMIX, есть дополнительная опция для случайного старта, чтобы уменьшить этот эффект путем случайного выбора подвыборки из данных, которая потом случайно приписывается *g* компонентам. Как описано в McLachlan and Peel (2000, Sect. 2.12), подвыборка должна быть достаточно большой, чтобы убедиться, что первый M-шаг может произвести недегенеративную оценку вектора параметров .

Ueda and Nakano (1998) предложили алгоритм детерминированного отжига EM (DAEM) для того, чтобы итерационный процесс EM мог восстановиться после плохого выбора начальных значений. Они предложили использовать принцип максимальной энтропии и аналог статистической механики, когда, параметр, скажем θ, вводится с соответствующей «температуре» в смысле отжига. С их алгоритмом DAEM. E-шаг осуществляется путем усреднения по распределению взятому пропорционально к текущей оценке условной плотности полных данных (имея наблюдаемые данные) в степени θ; см пример McLachlan and Peel (2000, pp. 58–60). В 2005 Pernkopf and Bouchaffra (2005) соединили генетический алгоритм (GA) и EM алгоритм для подгонки смесей нормальных распределений, где предложенный алгоритм менее чувствителен к начальному приближению и позволяет выбираться из локально оптимальных решений.

## Правила остановки

Необходимо иметь разумные критерии для остановки итерационного ЕМ-алгоритма. Существует несколько подходов:

Выберем малое число .

1. Условимся останавливать ЕМ-алгоритм, если расстояние между эмпирической функцией распределения и теоретической смесью функций распределения, в которую подставлены параметры, вычисленные с помощью ЕМ-алгоритма, становится меньше .

Где

- теоретическая смесь нормальных функций распределения, в которую подставлены параметры, вычисленные с помощью ЕМ-алгоритма на m-интеграции,

-эмпирическая функция распределения (, если , и в противном случае).

1. ЕМ-алгоритм можно останавливать, когда разность между значениями функции правдоподобия, вычисленными на соседних итерациях, мала, то есть дальнейшая работа ЕМ-алгоритма практически не увеличит текущее правдоподобие:
2. ЕМ-алгоритм можно останавливать, когда расстояние между оценками параметров, вычисленными на соседних итерациях мало, то есть дальнейшая работа ЕМ-алгоритма практически не изменит уже полученные оценки:

В соотношениях (3.3) и (3.5) нормы можно определять по-разному.

У всех перечисленных критериев остановки есть и достоинства, и недостатки.

Критерий (3.3) основанный на близости эмпирической функции распределения и теоретической смеси функции распределения, в которую подставлены параметры, вычисленные с помощью ЕМ-алгоритма, дает хорошую возможность судить о том насколько хороша полученная аппроксимация. Однако ЕМ-алгоритм может никогда не достичь таких значений параметров теоретической функции распределения, которые обеспечивают хорошее согласие с имеющимися наблюдениями.

Критерий (3.4), ориентирующийся на функцию правдоподобия естественно согласуется с методом, используемым в ЕМ-алгоритме, но как бы сводит все имеющиеся возможности, так сказать, степени свободы алгоритма, к единственному значению. Согласно этому критерию, алгоритм может быть остановлен если последовательно вычисленные значения *L* близки, но параметры далеки от точки максимума (функция правдоподобия может принимать близкие значения на 2х достаточно отдаленных друг от друга множествах значений параметров).

Критерий (3.5) использующий расстояние между последовательно полученными оценками вектора параметров, конечно же, остановит ЕМ-алгоритм, если последний достигнет точки максимума функции правдоподобия. К сожалению, этот критерий так же остановит алгоритм в случае, когда алгоритм достигнет области, далекой от точки максимума, но в которой скорость сходимости очень низка. [1]

# Свойства ЕМ Алгоритма

Плюсы:

1. Численная устойчивость. Алгоритм является абсолютно устойчивым, поскольку на каждой итерации происходит увеличение функции правдоподобия.
2. Монотонность.
3. Простота реализации. В частности, относительно просто программируется и не требует большого объема памяти. Отслеживая монотонную сходимость функции правдоподобия легко отслеживать сходимость и ошибки программирования McLachlan and Krishnan (1997, Sect. 1.7).
4. Цена на итерацию. Хотя ЕМ алгоритму и требуется больше итераций чем другим алгоритмам, затраты на итерацию ниже.
5. Может использоваться для получения оценок отсутствующих данных

Минусы:

1. Неустойчивость по начальным данным. Алгоритм сходится к локальному максимуму функции правдоподобия.
2. Иногда очень медленно сходится
3. В некоторых задачах Е- и М- шаги могут быть аналитически неразрешимы.

# Применение ЕМ алгоритма

## Разделение конечных смесей нормальных распределений

Определим функцию плотности вероятности смеси нормальных распределений как

Где это p-мерная нормальная функция плотности вероятности с математическим ожиданием и ковариационной матрицей . Вектор неизвестных параметров состоит из пропорций смеси , элементов мат ожидания и отдельных компонент ковариационной матрицы **.** Задача оценки это пример проблемы разделения смесей или в языке распознавания образов «задача обучения без учителя».

Рассмотрим соответствующую задачу «обучения с учителем», где наблюдения над случайным вектором **,** такие, что . Здесь является ненаблюдаемым компонентами-индикатора вектором, где *i*-ый элемент берется нулем или единицей, в соответствии с тем, является ли j-м наблюдение приходит из i-го компонента .Задача нахождения Оценок Максимального Правдоподобия гораздо проще в данном случае. Векторы классификаторы можно называть отсутствующими данными. Задача обучения без учителя так же может называться задачей с неполными данными, а задача обучения с учителем, задачей с полными данными. Относительно простой итерационный метод для вычисления Оценок Максимального Правдоподобия для задачи обучения без учителя может быть дан, используя простоту метода Оценок Максимального Правдоподобия для обучения с учителем. Это суть ЕМ алгоритма.

Логарифмическая функция правдоподобия на полных данных для θ определяется как

ЕМ алгоритм для этой задачи начинается с некоторого начального значения . Так как линейная функция ненаблюдаемых данных *z* для задачи, вычисление на Е шаге осуществляется просто заменой его текущим условным математическим Ожиданием с учетом наблюдаемых данных y, которые являются обычно апостериорной вероятностью *j*-го наблюдения из *i*-го компонента.

Из (5.2) следует что

Для смесей с нормальными плотностями, численно выгоднее работать в терминах достаточных статистик (Ng and McLachlan 2003) данных

Дифференцируя (5.4) по θ на базисе достаточных статистик в (5.5), М шаг представляется как

Е и М шаги повторяют до выполнения условия сходимости. В отличии от метода Максимального Правдоподобия для задачи с учителем, на М шаге задачи без учителя, используются апостериорные вероятности . Вектор математических ожиданий и ковариационная матрица вычисляются, используя как веса во взвешенных средних.

В случае неограниченной компонент-ковариационной матрицы , не ограничена, так как каждая точка данных дает прирост к сингулярности на границе параметрического пространства. (McLachlan and Peel 2000, Sect. 3.8) На практике, компонент-ковариационные матрицы может быть наложено ограничение, , где не определена. В этом случае гомоскедастических нормальных компонент, обновленная оценка общей компонентно-ковариационной матрицы определяется как  
где определяется (5.6), а обновление и также как выше в гетероскедастичном случае.

# Модификации ЕМ алгоритма

## Медианные модификации ЕМ-алгоритма

Как было экспериментально установлено, ЕМ-алгоритм обладает сильной неустойчивостью по начальным данным. Например, в случае четырехкомпонентной смеси нормальных законов при объеме выборки 200–300 наблюдений замена лишь одного наблюдения другим может кардинально изменить итоговые оценки, полученные с помощью ЕМ-алгоритма. Возможно, эта неустойчивость обусловлена тем, что стандартные (наиболее правдоподобные для случая нормального распределения) оценки математического ожидания и дисперсии (среднее арифметическое и выборочная дисперсия) при “засорении” (контаминации) выборки “посторонними” или “паразитными” наблюдениями становятся заметно менее эффективными по сравнению со, скажем, выборочной медианой. Этот эффект обнаружен Дж. Тьюки (Tukey, 1960) и описан, например, в (Айвазян и др., 1983) и (Королев, 2006). Формально модель контаминации Тьюки сводится к тому, что вместо “чистого” модельного распределения, интерпретируемого как однородная модель, в качестве модельного распределения рассматривается неоднородная модель, имеющая вид смеси исходного “чистого” распределения и некоторого другого закона, описывающего “засоряющие” наблюдения. В задаче разделения смесей по самой сути модели, когда оцениваются параметры одной компоненты смеси, наблюдения с распределениями, соответствующими другим компонентам, являются «загрязняющими». Это обстоятельство может сыграть особенно важную роль при реализации SEM-алгоритма, описываемого ниже.

Для противодействия указанной неустойчивости ЕМ-алгоритма можно использовать медианные модификации ЕМ-алгоритма. Смысл этих модификаций в том, что наиболее “неустойчивые” этапы выполнения ЕМ-алгоритма заменяются более устойчивыми. В частности, на М-этапе неустойчивые моментные оценки наибольшего правдоподобия (которые для нормальных компонент минимизируют квадратичный риск) заменяются более устойчивыми (робастными) оценками медианного типа, оптимальными в смысле среднего абсолютного отклонения.

Далее будут приведены конечные результаты для смеси нормальных распределений. Математическое обоснование данного метода можно посмотреть в [1].

## Первая медианная модификация

Введем следующие величины:

имеющие смысл некой вероятности. Введем также «фиктивные» случайные величины принимающие значения с вероятностями . Переупорядочим значения случайной величины по неубыванию, одновременно переставляя соответствующие данным значениям вероятности. Пусть — вероятность, соответствующая значению. Положим:

Тогда

Пусть тогда

## Вторая медианная модификация

Матожидание оценивается так же, как и в первой медианной модификации. Рассмотрим дисперсию .

Обозначим

Тогда

## Stochastic EM(SEM) алгоритм

Классический ЕМ-алгоритм относится к категории так называемых «жадных» алгоритмов (greedy algorithms) в том смысле, что он «бросается» на первый попавшийся локальный максимум. Другими словами, будучи методом *локальной оптимизации*, он приводит не к глобальному максимуму функции правдоподобия, а к тому локальному максимуму, который, грубо говоря, является ближайшим к начальному приближению.

Самый простой способ противодействия этому свойству заключается в том, чтобы, не ограничиваясь единственным начальным приближением и, соответственно единственной траекторией ЕМ-алгоритма, реализовать несколько траекторий, задавая (например, случайно) нескольких различных начальных приближений, а затем выбрать тот из результатов, для которого правдоподобие является наибольшим среди всех реализованных траекторий ЕМ-алгоритма. Однако при таком подходе остаётся неясным ответ о том, каким механизмом разумнее всего пользоваться при переходе от одного начального приближения к другому. В частности, когда начальное приближение задается случайно, без дополнительной информации нельзя исчерпывающим образом определить распределение вероятностей, в соответствии с которым следует генерировать очередное начальное приближение.

Другой, оказавшийся весьма эффективным, способ заключается как бы в случайном, но целенаправленном «встряхивании» наблюдений (выборки) на каждой итерации. Этот способ лежит в основе SEM-алгоритма, название которого является аббревиатурой термина Stochastic EM-algorithm (Стохастический (или случайный) ЕМ-алгоритм). SEM-алгоритм, предложенный в работах (Broniatowski, Celeux and Diebolt, 1983), (Celeux and Diebolt, 1984), (Celeux and Diebolt, 1985), весьма прост.

* SEM работает относительно быстро, и его результаты практически не зависят от начального приближения;
* как правило, SEM находит экстремум , близкий к глобальному.

Пусть вся выборка разбита на кластеры : каждый элемент относится к единственному(!) кластеру , то есть утверждается, что данный элемент взят из компоненты смеси.

S-шаг

На первом этапе SEM-алгоритма производится так называемое стохастическое моделирование. Для каждого , m генерируется вектор как реализация случайного вектора из полиномиального распределения с параметрами 1 и , где — это вероятность того, что величина равна 1. По векторам определяется разбиение выборки на кластеры и соответствующие численности кластеров .

Е-шаг

Остается без изменений.

М-шаг

Пересчитываются веса:

и вместо максимизации взвешенного правдоподобия:

решается задача обычного невзвешенного правдоподобия:

Для SEM-алгоритма также существуют медианные модификации. Ознакомиться с ними можно в [1].

## Классификационный ЕМ-алгоритм (Classification EM, CEM)

Этот алгоритм совпадает с SEM-алгоритмом, за исключением того, что вместо S-шага используется детерминированное правило, эквивалентное классификации по принципу максимума апостериорной вероятности, то есть приписывается тому кластеру, номер которого совпадает с номером наибольшего из чисел . Формулы для оценок параметров аналогичны SEM-алгоритму.

## Обобщённый ЕМ-алгоритм (Generalized EM, GEM)

В тех случаях, когда максимизация функционала , имеющего смысл полного правдоподобия, по каким-либо причинам затруднена, применяется подход, в котором достаточно лишь сместиться в сторону максимума значения функционала, сделав одну или несколько итераций на М-шаге. Этот алгоритм также обладает неплохой сходимостью.

# Заключение

Подводя итоги, ЕМ алгоритм представляет собой мощный инструмент прикладной статистики. В данной работе рассмотрены основы алгоритма и основные его модификации, а также одно из применений – разделение смеси нормальных компонент.

# Список Литературы

1. *Королев В. Ю.* Вероятностно-статистические методы декомпозиции волатильности хаотических процессов. — Издательство Московского университета Москва, 2011. — 512 с.
2. *Gentle James et al.* Handbook of Computational Statistics: Concepts and Methods 2nd edition. — Springer, 2012 — 1192 pp.
3. *Martin Haugh*: Machine Learning for OR&FE, 2015- 7 p
4. *McLachlan, Geoffrey J*., The EM algorithm and extensions / GJ. McLachlan and T. Krishnan. p.- (Wiley series in probability and statistics. Applied probability and statistics), 1997

В этом разделе мы убедимся, что, если положить , специальным образом выбрать штрафную функцию и положить , то описанный выше ЕМ-алгоритм будет алгоритмом проксимального типа.

Пусть жто условная плотность при заданном . Тогда полная функция правдоподобия может быть выражена как

Беря мат.ожидание с обоих сторон получаем