EM-алгоритм

1 Мотивация

Пусть

$$\boldsymbol{\eta} \sim \mathrm{Mult}_4(n, \mathbf{p}), \quad \mathbf{p} = \left(\frac{1}{2} + \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right)$$

с плотностью

$$q_\theta(\mathbf{y}) = \frac{n!}{y_1! y_2! y_3! y_4!} \left(\frac{1}{2} + \frac{\theta}{4}\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4}.$$

Выборка $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$; по ОМП,

$$\begin{aligned}
\log q_\theta(\mathbf{Y}) &\propto \sum_{i=1}^N y_1^{(i)} \log\left(\frac{1}{2} + \frac{\theta}{4}\right) + y_2^{(i)} \log\frac{1-\theta}{4} + y_3^{(i)} \log\frac{1-\theta}{4} + y_4^{(i)} \log\frac{\theta}{4} \\
\frac{\partial \log q_\theta(\mathbf{Y})}{\partial\theta} &\propto \sum_{i=1}^N y_1^{(i)} \frac{1/4}{1/2 + \theta/4} - (y_2^{(i)} + y_3^{(i)}) \frac{1}{1-\theta} + y_4^{(i)} \frac{1}{\theta} \\
&= \sum_{i=1}^N \frac{y_1^{(i)}}{2+\theta} + \frac{y_2^{(i)} + y_3^{(i)}}{\theta - 1} + \frac{y_4^{(i)}}{\theta}.
\end{aligned}$$

Получили кубическое[1] уравнение относительно θ (могло быть и выше).

С другой стороны, пусть

$$\boldsymbol{\xi} \sim \mathrm{Mult}_5(n, \mathbf{p}), \quad \mathbf{p} = \left(\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right).$$

ОМП для $\boldsymbol{\xi}$:

$$\begin{aligned}
p_\theta(\mathbf{x}) &\propto \left(\frac{1}{2}\right)^{x_1} \left(\frac{\theta}{4}\right)^{x_2} \left(\frac{1-\theta}{4}\right)^{x_3+x_4} \left(\frac{\theta}{4}\right)^{x_5} \\
\log p_\theta(\mathbf{x}) &\propto x_1 \log\frac{1}{2} + x_2 \log\frac{\theta}{4} + (x_3 + x_4) \log\frac{1-\theta}{4} + x_5 \log\frac{\theta}{4} \\
\frac{\partial \log p_\theta}{\partial\theta} &\propto \frac{x_2 + x_5}{\theta} - \frac{x_3 + x_4}{1-\theta} = 0 \iff \hat{\theta} = \frac{x_2 + x_5}{x_2 + x_3 + x_4 + x_5}.
\end{aligned}$$

Так что $\hat{\theta}$ находится достаточно просто.

Исходную задачу оценки параметров $\boldsymbol{\eta}$ можно переформулировать через оценку параметров $\boldsymbol{\xi}$:

$$\boldsymbol{\eta} = (\xi_1 + \xi_2, \xi_3, \xi_4, \xi_5).$$

В примерах выше тогда можно думать про $\mathbf{x}$ как про полные данные, $\mathbf{y}$ — неполные данные (с «пропусками»), причем

$$y_1 = x_1 + x_2,\ y_2 = x_3,\ y_3 = x_4,\ y_4 = x_5.$$

[1] На самом деле квадратное.

А для полных данных функция правдоподобия, как было видно, будет иметь более простой вид.

Однако элемент выборки x_2 не наблюдается. Можно тогда посчитать мат. ожидание:

$$\begin{aligned}\mathsf{E}(\boldsymbol{\xi} \mid \boldsymbol{\eta} = \mathbf{y}) &= (\mathsf{E}(\xi_1 \mid \xi_1 + \xi_2 = y_1), \mathsf{E}(\xi_2 \mid \xi_1 + \xi_2 = y_1), y_2, y_3, y_4) = \\ &= \left(y_1 \frac{1/2}{1/2 + \theta/4}, y_1 \frac{\theta/4}{1/2 + \theta/4}, y_2, y_3, y_4 \right).\end{aligned}$$

Но

$$\mathsf{P}(\xi_1 = x \mid \xi_1 + \xi_2 = y_1) = \frac{\mathsf{P}(\xi_1 = x, \xi_2 = y_1 - x)}{\mathsf{P}(\xi_2 = y_1 - x)} =$$

Пусть есть приближение $\hat{\theta}^{(n)}$. Тогда

$$\hat{x}_2^{(n)} = \frac{\hat{\theta}^{(n)}/4}{1/2 + \hat{\theta}^{(n)}/4} y_1$$

и ОМП по полной выборке есть

$$\hat{\theta}^{(n+1)} = \frac{\hat{x}_2^{(n)} + x_5}{\hat{x}_2^{(n)} + x_3 + x_4 + x_5}.$$

Необходимо удостовериться, что оценка сойдется куда нужно.

2 EM-алгоритм

2.1 Формулировка

Пусть $\eta \sim \mathcal{Q}(\theta)$ на $(\mathfrak{Y}, \mathcal{B})$ с плотностью q_θ, $\theta \in \Theta$; $\mathbf{y}$ — *неполные* данные. Пусть $\xi \sim \mathcal{P}(\theta)$ на $(\mathfrak{X}, \mathcal{A})$ с плотностью p_θ такой, что относительно просто по *дополненной* выборке $\mathbf{x}$ можем посчитать ФП, так что

$$\hat{\theta} = \operatorname*{argmax}_{\theta \in \Theta} \log p_\theta(\mathbf{x}),$$

причем

$$\eta = T(\xi).$$

Наблюдается η, а ξ — нет.

Алгоритм. *Пусть $\hat{\theta}^{(k)}$ — текущее приближение параметра.*

1. *Expectation:*
$$Q^{(k)}(\theta) = \mathsf{E}_{\hat{\theta}^{(k)}}(\log p_\theta(\mathbf{x}) \mid T(\mathbf{x}) = \mathbf{y}).$$

2. *Maximization:*
$$\hat{\theta}^{(k+1)} = \operatorname*{argmax}_{\theta} Q^{(k)}(\theta).$$

Замечание. $Q(\theta)$ в некоторых задачах считается относительно хорошо. Кроме того, можем сами выбирать T (обычно проекция) и распределение полных данных ξ. Фиксировано только η.

2.2 EM алгоритм для распределений из экспоненциального семейства

Определение. *Экспоненциальное семейство* есть семейство плотностей вида

$$p_\theta(x) = \frac{b(x)}{a(\theta)} \exp\left\{ \mathbf{c}^\mathsf{T}(\theta) \mathbf{t}(x) \right\},$$

$\mathbf{t}$ — достаточные статистики.

Для таких плотностей

$$
\begin{aligned}
Q^{(k)}(\theta) &= \mathsf{E}_{\hat\theta^{(k)}}\Big(\log b(x) - \log a(\theta) + \mathbf{c}^{\mathsf{T}}(\theta)\mathbf{t}(x) \mid \boldsymbol\eta = \mathbf{y}\Big) \\
&= \mathsf{E}_{\hat\theta^{(k)}}(\log b(x) \mid \boldsymbol\eta = \mathbf{y}) - \log a(\theta) + \mathbf{c}^{\mathsf{T}}(\theta)\mathsf{E}_{\hat\theta^{(k)}}(\mathbf{t}(x) \mid \boldsymbol\eta = \mathbf{y}),
\end{aligned}
$$

так что всё, что нужно уметь делать — считать

$$
\mathsf{E}_{\hat\theta^{(k)}}(\mathbf{t}(x) \mid T(\mathbf{x}) = \mathbf{y}),
$$

тем более, что $\mathbf{t}(x)$ часто — суммы или суммы квадратов.

2.3 Свойства алгоритма

- Не предполагая ничего дополнительно, можем доказать, что последовательность $\hat\theta^{(k)}$ приводит к неуменьшению $\log p_\theta(\mathbf{x}) = \mathsf{L}(\theta)$:

  $$
  \mathsf{L}(\hat\theta^{(k)}) \ge \mathsf{L}(\hat\theta^{(k-1)})
  $$

  по неравенству Єнсена.

- Если требовать дополнительно, например, регулярность, то

  $$
  \mathsf{L}(\hat\theta^{(k)}) > \mathsf{L}(\hat\theta^{(k-1)})
  $$

  Но если разница пропорциональна $1/k$, то ни к чему не сойдемся — будем делать бесконечно мелкие шаги. Может застрять в локальном максимуме или на плато.

…ЕМ-алгоритм в каждой точке функции правдоподобия строит наилучшие приближения…

Замечание. …ММ-алгоритм (требование выполнения неравенства Єнсена. для ϕ)…

3 Задача о разделении смеси

3.1 ЕМ для нормальной смеси

3.1.1 Две компоненты

Пусть дана выборка $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$. Зададим модель как смесь двух нормальных величин $\boldsymbol\eta^{(1)} \sim \mathrm{N}(\boldsymbol\mu^{(1)}, \boldsymbol\Sigma^{(1)})$, $\boldsymbol\eta^{(2)} \sim \mathrm{N}(\boldsymbol\mu^{(2)}, \boldsymbol\Sigma^{(2)})$:

$$
\boldsymbol\eta = \boldsymbol\eta^{(\zeta+1)} = (1-\zeta)\cdot\boldsymbol\eta^{(1)} + \zeta\cdot\boldsymbol\eta^{(2)}, \quad \zeta \in \{0,1\}, \mathsf{P}(\zeta) = p.
$$

Тогда плотность $\boldsymbol\eta$ есть

$$
q(\mathbf{y}) = (1-p)\phi^{(1)}(\mathbf{y}) + p\phi^{(2)}(\mathbf{y}), \quad \phi^{(\ell)} = \mathrm{pdf}_{\mathrm{N}(\boldsymbol\mu^{(\ell)}, \boldsymbol\Sigma^{(\ell)})}
$$

Так что нужно оценить параметры $\boldsymbol\theta = (p, \boldsymbol\mu^{(1)}, \boldsymbol\mu^{(2)}, \boldsymbol\Sigma^{(1)}, \boldsymbol\Sigma^{(2)})$. Логарифм ФП

$$
\log \mathsf{L}_q(\boldsymbol\theta) = \sum_{i=1}^{N} \log\Big[(1-p)\phi^{(1)}(\mathbf{y}_i) + p\phi^{(2)}(\mathbf{y}_i)\Big]
$$

может быть сложно максимизировать напрямую.

Пусть для каждой точки $\mathbf{y}_j$ известно, из какой она компоненты, т.е. заданы пары $((\mathbf{y}_j, z_j))_{j=1}^N = \mathbf{X}$. На этих (полных) данных, плотность есть

$$
p(\mathbf{x}) = \Big[(1-p)\phi^{(1)}(\mathbf{y})\Big]^{1-z}\Big[p\phi^{(2)}(\mathbf{y})\Big]^{z},
$$

и логарифм (полной) ФП

$$
\begin{aligned}
\log \mathsf{L}_p(\theta) &= \log \prod_{j=1}^{N} \left[(1-p)\phi^{(1)}(\mathbf{y}_j)\right]^{1-z_j} \left[p\phi^{(2)}(\mathbf{y}_j)\right]^{z_j} \\
&= \sum_{j=1}^{N} (1-z_j) \log \left[(1-p)\phi^{(1)}(\mathbf{y}_j)\right] + z_j \log \left[p\phi^{(2)}(\mathbf{y}_j)\right] \\
&= \sum_{j=1}^{N} (1-z_j) \log \phi^{(1)}(\mathbf{y}_j) + z_j \log \phi^{(2)}(\mathbf{y}_j) + \sum_{j=1}^{N} (1-z_j) \log(1-p) + z_j \log p.
\end{aligned}
$$

Находя максимум, можно показать, что

$$
\hat{\boldsymbol{\mu}}^{(\ell)} = \bar{\mathbf{Y}}^{(\ell)}, \quad \hat{\mathbf{\Sigma}}^{(\ell)} = \widehat{\operatorname{cov}} \mathbf{Y}^{(\ell)}, \quad \hat{p} = \sum_{j=1}^{N} z_j / N.
$$

где $\mathbf{Y}^{(\ell)}$ — часть данных, для которых ℓ-я компонента не нулевая.

Вместо неизвестных z_j подставим

$$
\begin{aligned}
\zeta_j(\boldsymbol{\theta}) &= \mathsf{E}_{\hat{\boldsymbol{\theta}}_k}(z_j \mid \mathbf{Y}) = \mathsf{P}_{\hat{\boldsymbol{\theta}}_k}(z_j = 1 \mid \eta_j = \mathbf{y}_j) = \frac{\mathsf{P}(\eta_j = \mathbf{y}_j \mid z_j = 1)\mathsf{P}(z_j = 1)}{\mathsf{P}(\eta_j = \mathbf{y}_j)} \\
&= \frac{\mathsf{P}(\eta_j = \mathbf{y}_j \mid z_j = 1)\mathsf{P}(z_j = 1)}{\mathsf{P}(\eta_j = \mathbf{y}_j \mid z_j = 0)\mathsf{P}(z_j = 0) + \mathsf{P}(\eta_j = \mathbf{y}_j \mid z_j = 1)\mathsf{P}(z_j = 1)} \\
&= \frac{\phi^{(2)}(\mathbf{y}_j)p}{\phi^{(1)}(\mathbf{y}_j)(1-p) + \phi^{(2)}(\mathbf{y}_j)p}, \quad j \in 1:N.
\end{aligned}
$$

Тогда в обозначении

$$
w_j^{(1)} := \frac{1-\hat{\zeta}_j}{\sum_{j'=1}^{N}(1-\hat{\zeta}_{j'})} \qquad w_j^{(2)} := \frac{\hat{\zeta}_j}{\sum_{j'=1}^{N}\hat{\zeta}_{j'}}
$$

оценки примут вид

$$
\hat{\mu}_i^{(\ell)} = \sum_{j=1}^{N} w_j^{(\ell)} y_{ji} \qquad \widehat{\operatorname{cov}}\left(\eta_{i_1}^{(\ell)}, \eta_{i_2}^{(\ell)}\right) = \sum_{j=1}^{N} w_j^{(\ell)} (y_{ji_1} - \hat{\mu}_{i_1}^{(\ell)})(y_{ji_2} - \hat{\mu}_{i_2}^{(\ell)}) \qquad \hat{p} = \sum_{i=1}^{N} \hat{\zeta}_i / N
$$

Векторизованная форма Пусть

$$
\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 & \ldots & \mathbf{y}_N \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} y_{11} & \cdots & y_{1d} \\ \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{Nd} \end{pmatrix}, \quad \hat{\boldsymbol{\zeta}} = \begin{pmatrix} \hat{\zeta}_1 \\ \vdots \\ \hat{\zeta}_N \end{pmatrix}.
$$

Тогда, в обозначении

$$
\mathbf{w}^{(1)} = \frac{(\mathbf{1} - \hat{\boldsymbol{\zeta}})}{\mathbf{1}^{\mathsf{T}}(\mathbf{1} - \hat{\boldsymbol{\zeta}})} \qquad \mathbf{w}^{(2)} = \frac{\hat{\boldsymbol{\zeta}}}{\mathbf{1}^{\mathsf{T}}\hat{\boldsymbol{\zeta}}}
$$

оценки примут вид

$$
\hat{\boldsymbol{\mu}}^{(\ell)} = \begin{pmatrix} \sum_{j=1}^{N} w_j^{(\ell)} y_{j1} \\ \vdots \\ \sum_{j=1}^{N} w_j^{(\ell)} y_{jd} \end{pmatrix} = \mathbf{Y}^{\mathsf{T}} \mathbf{w}^{(\ell)} \qquad \hat{\mathbf{\Sigma}}^{(\ell)} = \ldots.
$$

Алгоритм. *Выбрать начальные значение* $\hat{\boldsymbol{\mu}}^{(\ell)} = \mathbf{y}_{j_\ell}, j_\ell \in 1:N$, $\hat{\mathbf{\Sigma}}^{(\ell)}$, $\hat{p} = 1/2$.

1. *Expectation: Найти* $\hat{\zeta}_j, \quad j \in 1:N$.

2. *Maximization: по формулам выше пересчитать* $\hat{\boldsymbol{\mu}}^{(\ell)}, \hat{\mathbf{\Sigma}}^{(\ell)}, \hat{p}$.

3.1.2 Произвольное количество компонент

Пусть

$$\boldsymbol{\eta} = \boldsymbol{\eta}^{(\zeta)}, \quad \boldsymbol{\eta}^{(\ell)} \sim \mathrm{N}(\boldsymbol{\mu}^{(\ell)}, \boldsymbol{\Sigma}^{(\ell)}),\ \zeta \in \{1, \dots, m\}, \mathsf{P}(\zeta = \ell) = p_\ell.$$

- Плотность $\boldsymbol{\eta}$ есть

$$q(\mathbf{y}) = \sum_{\ell=1}^{m} p_\ell \phi^{(\ell)}(\mathbf{y}), \quad \phi^{(\ell)} = \mathrm{pdf}_{\mathrm{N}(\boldsymbol{\mu}^{(\ell)}, \boldsymbol{\Sigma}^{(\ell)})}.$$

- Логарифм ФП $\boldsymbol{\eta}$

$$\log \mathsf{L}_q(\boldsymbol{\theta}) = \sum_{j=1}^{N} \log \sum_{\ell=1}^{m} p_\ell \phi^{(\ell)}(\mathbf{y}_j).$$

- Плотность $\boldsymbol{\xi}$:
- Логарифм (полной) ФП
- Оценка z_j:

$$\zeta_j^{(\ell)}(\boldsymbol{\theta}) = \mathsf{P}_{\hat{\boldsymbol{\theta}}_k}(z_j = \ell \mid \eta_j = \mathbf{y}_j) = \frac{\phi^{(\ell)}(\mathbf{y}_j) p_\ell}{\sum_{\ell'=1}^{m} \phi^{(\ell')}(\mathbf{y}_j) p_{\ell'}}, \quad j \in 1:N.$$

- Веса:

$$w_j^{(\ell)} := \frac{\zeta_j^{(\ell)}}{\sum_{j'=1}^{N} \zeta_{j'}^{(\ell)}}.$$

- Оценки:

$$\hat{\mu}_i^{(\ell)} = \sum_{j=1}^{N} w_j^{(\ell)} y_{ji} \qquad \widehat{\mathrm{cov}}\left(\eta_{i_1}^{(\ell)}, \eta_{i_2}^{(\ell)}\right) = \sum_{j=1}^{N} w_j^{(\ell)} (y_{ji_1} - \hat{\mu}_{i_1}^{(\ell)})(y_{ji_2} - \hat{\mu}_{i_2}^{(\ell)}) \qquad \hat{p}_\ell = \frac{1}{N} \sum_{j=1}^{N} \hat{\zeta}_j^{(\ell)}$$

Векторизованная форма Пусть

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 & \dots & \mathbf{y}_N \end{pmatrix}^{\mathsf{T}} = \begin{pmatrix} y_{11} & \dots & y_{1d} \\ \vdots & \ddots & \vdots \\ y_{N1} & \dots & y_{Nd} \end{pmatrix}, \quad \hat{\boldsymbol{\zeta}}^{(\ell)} = \begin{pmatrix} \hat{\zeta}_1^{(\ell)} \\ \vdots \\ \hat{\zeta}_N^{(\ell)} \end{pmatrix}.$$

Тогда, в обозначении

$$\mathbf{w}^{(\ell)} = \frac{\hat{\boldsymbol{\zeta}}^{(\ell)}}{\mathbf{1}^{\mathsf{T}} \hat{\boldsymbol{\zeta}}^{(\ell)}}$$

оценки примут вид

$$\hat{\boldsymbol{\mu}}^{(\ell)} = \begin{pmatrix} \sum_{j=1}^{N} w_j^{(\ell)} y_{j1} \\ \vdots \\ \sum_{j=1}^{N} w_j^{(\ell)} y_{jd} \end{pmatrix} = \mathbf{Y}^{\mathsf{T}} \mathbf{w}^{(\ell)} \qquad\qquad \hat{\boldsymbol{\Sigma}}^{(\ell)} = \dots.$$

Алгоритм. *Выбрать начальные значение* $\hat{\boldsymbol{\mu}}^{(\ell)} = \mathbf{y}_{j_\ell}, j_\ell \in 1:N$, $\hat{\boldsymbol{\Sigma}}^{(\ell)}$, $\hat{\mathbf{p}} = 1/m$.

1. *Expectation: Найти* $\hat{\boldsymbol{\zeta}}^{(\ell)}, \quad \ell \in 1:m$.
2. *Maximization: по формулам выше пересчитать* $\hat{\boldsymbol{\mu}}^{(\ell)}, \hat{\boldsymbol{\Sigma}}^{(\ell)}, \hat{\mathbf{p}}$.

Замечание. z_j можно проинтерпретировать как апостериорные оценки принадлежности к той или иной компоненты смеси. …Bayesian / frequentist…

Замечание. Пусть $\dim \boldsymbol{\eta} = d$, количество компонент m; тогда количество параметров

$$(m-1) + md + m \cdot \frac{d(d+1)}{2}.$$

Замечание. Хуже всего сходится ковариационная матрица (должна быть оценена вся сразу). В качестве индикатора сходимости используют либо $-\log \mathsf{L}$ либо сходимость ковариационной матрицы в себе.

Замечание. $\max_{\boldsymbol{\theta}} \log \mathsf{L} = \infty$ достигается в точке пространства параметров, где $\boldsymbol{\mu}_1 = \mathbf{y}_i$, $\boldsymbol{\Sigma} = \mathbf{0}$, однако «хорошим» решением это не является, поэтому ищут лишь подходящий локальный максимум. Из всевозможных локальных максимумов выбирают такой (запуская алгоритм несколько раз с разными начальными параметрами), что на нем величина $\log \mathsf{L}$ наибольшая.

Замечание. У q 2 глобальных максимума, так что оценки МП не вполне правомочны. В общем случае m компонент, экстремумов $m!$.

3.2 Model-based Clustering

Пусть количество компонент m произвольно. Задача:

- Выбрать m.
- Выбрать подходящую структуру зависимости данных — $\boldsymbol{\Sigma}$.
- Оценить параметры..

Структура $\boldsymbol{\Sigma}$ Положительно определенную $\boldsymbol{\Sigma}$ можно «привести к главным осям»

$$\boldsymbol{\Sigma} = \lambda \mathbf{D}^{\mathsf{T}} \mathbf{A} \mathbf{D},$$

где $\mathbf{D}$ — ортогональная матрица поворота, $\operatorname{tr} \mathbf{A} = 1$. Варианты

- $\boldsymbol{\Sigma} = \lambda \mathbf{I}$, т.е. все компоненты независимы и с одинаковой дисперсии, так что параметр вообще один, либо
- $\boldsymbol{\Sigma} = \lambda \mathbf{A}$ — компоненты некоррелированы, но разная дисперсия, либо
- $\boldsymbol{\Sigma} = \lambda_k \mathbf{I}$ (своя дисперсия в каждой компоненте), либо
- $\boldsymbol{\Sigma} = \lambda_k \mathbf{A}_k$ и т.д.

Замечание. В R — `Mclust`. Позволяет оценить параметры, когда выборка из смеси нормальных распределений, задать модели для ковариационных матриц и выбрать наилучшую. Любая модель задается аббревиатурой

`I` Identity

`E` Equal

`V` Variate

Как признак меняется по разным компонентам смеси. Можно думать об объеме (λ), форме ($\mathbf{A}$) и ориентации ($\mathbf{D}$).

Пример. $\boldsymbol{\Sigma}_k = \lambda \mathbf{I}$ соответствует `EII`.

4 Information Criteria

Чем более общая модель, тем выше $\log \mathsf{L}$. Напрямую сравнивать нельзя, но можем вычитать штраф за количество параметров $f(\mathrm{df})$; тогда

$$\log \mathsf{L}(\theta) - f(\mathrm{df}).$$

Определение. $\mathrm{AIC} = \log \mathsf{L}(\theta) - \mathrm{df}$.

Замечание. AIC работает только в случае, когда истинная модель содержится в пространстве параметров.

Определение. $\mathrm{BIC} = \log \mathsf{L}(\theta) - \mathrm{df}/2 \cdot \log N$.