Методы и алгоритмы анализа скрытых марковских моделей

Выпускная квалификационная работа

Зотиков Дмитрий Юрьевич

Научный руководитель: д.ф-м.н., проф. Кривулин Н.К.

Рецензент: к.ф-м.н., доцент Шпилев П.В.

Санкт-Петербургский государственный университет

Прикладная математика и информатика

Вычислительная стохастика и статистические модели

СтатМод

Задачи дипломной работы

Рассматриваются скрытые марковские модели (СММ) как зависимая смесь распределений для анализа временных рядов (Zucchini et al., 2016). Задачи:

- изучение предложенного формализма:
  - матричное представление алгоритмов;
  - произвольные плотности для процесса наблюдений;
  - численная максимизация функции правдоподобия;
  - вывод и предсказание наблюдений;
  - оптимальный выбор компонент смеси через информационные критерии.
- создание объектно-ориентированной реализации на R;
- тестирование алгоритма оценки параметров на модельных данных для разных распределений компонент смеси;
- изучение работоспособности алгоритмов интер- и экстраполяции.

Однородная СММ 1-го порядка

- $\{\kappa_t\}_{t\in\mathbb{N}}$ — ОМЦ 1-го порядка с пространством состояний $\{1,\ldots,m\}$, матрицей переходных вероятностей

$$\mathbf{A} = (a_{ij})_{i,j=1}^{m}, \quad a_{ij} = \mathsf{P}(\kappa_{t+1} = j \mid \kappa_t = i)$$

и вектором начального распределения

$$\mathbf{u} = (\mathsf{P}(\kappa_1 = 1), \ldots, \mathsf{P}(\kappa_1 = m))\,.$$

- $\{\xi_t\}_{t\in\mathbb{N}}$ — процесс наблюдений с распределением

$$\xi_t \sim \mathcal{P}_j(\boldsymbol{\theta}_j),\ \boldsymbol{\theta}_j = \left(\theta_1, \ldots, \theta_{r_j}\right) \quad \text{при } \kappa_t = j,\ j \in 1:m.$$

Если $p_j(x \mid \boldsymbol{\theta}_j)$ — плотность / масса $\mathcal{P}_j(\boldsymbol{\theta}_j)$, то

$$\mathbf{B}(x) = \operatorname{diag}\left(p_1(x \mid \boldsymbol{\theta}_1), \ldots, p_m(x \mid \boldsymbol{\theta}_m)\right).$$

- Выполняется свойство

$$\mathsf{P}(\xi_t \mid \boldsymbol{\xi}_{1:(t-1)}, \kappa_{1:t}) = \mathsf{P}(\xi_t \mid \kappa_t).$$

$\mathcal{M} = (\mathbf{A}, \mathbf{B}, \mathbf{u})$. Обозначим $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)$. Пусть дан ряд $\mathbf{x} = (x_1, \ldots, x_T)$.

- Можно показать (Zucchini et al., 2016), что

$$\begin{aligned} \mathsf{L}(\mathcal{M} \mid \mathbf{x}) &= \mathsf{P}(\boldsymbol{\xi} = \mathbf{x} \mid \mathcal{M}) = \mathbf{u}\mathbf{B}(x_1)\left(\prod_{t=2}^{T} \mathbf{A}\mathbf{B}(x_t)\right)\mathbf{1}^\mathsf{T} \\ &= \boldsymbol{\alpha}_T(\mathbf{x} \mid \mathcal{M})\mathbf{1}^\mathsf{T}, \quad \mathbf{1} = \begin{pmatrix} 1 & \ldots & 1 \end{pmatrix}. \end{aligned}$$

где

$$\boldsymbol{\alpha}_t(\mathbf{x} \mid \mathcal{M}) := \mathbf{u}\mathbf{B}(x_1)\left(\prod_{s=2}^{t} \mathbf{A}\mathbf{B}(x_s)\right), \quad \boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1}\mathbf{A}\mathbf{B}(x_t).$$

- Считают для нормализованных на каждом t

$$\bar{\boldsymbol{\alpha}}_t := \frac{\boldsymbol{\alpha}_t}{\boldsymbol{\alpha}_t \mathbf{1}^\mathsf{T}}.$$

- Традиционно (Rabiner, 89) — метод максимизации мат. ожидания (EM, «Баум–Уэлша»).
- Общие методы численной максимизации: градиентные (градиентного спуска, сопряженных градиентов, BFGS,...), неградиентные (Нелдера–Мида, имитации отжига, ...).
- (Turner, 2008): можно вычислить аналитический градиент и гессиан ФП; оптимизировать методом Левенберга–Марквардта (LM). 95% доверительные интервалы для среднего отношения времени работы:
  - EM к LM: $[6.49, 7.23]$.
  - BFGS к LM (без аналитического градиента): $[19.7, 21.35]$.
  - BFGS к LM (с аналитическим градиентом): $[2.9, 3.14]$.

$$
\begin{aligned}
&\mathsf{P}(\xi_t = x \mid \boldsymbol{\xi}_{1:T\setminus t} = \mathbf{x}_{1:T\setminus t}) \\
&\qquad = \frac{\mathsf{P}(\boldsymbol{\xi}_{1:t-1} = \mathbf{x}_{1:t-1}, \xi_t = x, \boldsymbol{\xi}_{t+1:T} = \mathbf{x}_{t+1:T})}{\mathsf{P}(\boldsymbol{\xi}_{1:t-1} = \mathbf{x}_{1:t-1}, \boldsymbol{\xi}_{t+1:T} = \mathbf{x}_{t+1:T})} \\
&= \frac{\mathbf{uB}(x_1)\mathbf{AB}(x_2)\ldots\mathbf{AB}(x_{t-1})\mathbf{AB}(x)\mathbf{AB}(x_{t+1})\ldots\mathbf{AB}(x_T)\mathbf{1}^\mathsf{T}}{\mathbf{uB}(x_1)\ldots\mathbf{AB}(x_{t-1})\mathbf{AIAB}(x_{t+1})\ldots\mathbf{AB}(x_T)\mathbf{1}^\mathsf{T}}.
\end{aligned}
$$

Обозначив

$$
\begin{aligned}
\boldsymbol{\alpha}_{t-1} &:= \mathbf{uB}(x_1)\mathbf{AB}(x_2)\ldots\mathbf{AB}(x_{t-1}) \\
\boldsymbol{\beta}_t^\mathsf{T} &:= \mathbf{AB}(x_{t+1})\ldots\mathbf{AB}(x_T)\mathbf{1}^\mathsf{T}
\end{aligned}
$$

получают

$$
\mathsf{P}(\xi_t = x \mid \boldsymbol{\xi}_{1:T\setminus t} = \mathbf{x}_{1:T\setminus t}) = \frac{\boldsymbol{\alpha}_{t-1}\mathbf{AB}(x)\boldsymbol{\beta}_t^\mathsf{T}}{\boldsymbol{\alpha}_{t-1}\mathbf{A}\boldsymbol{\beta}_t^\mathsf{T}}.
$$

$$
\begin{aligned}
\mathsf{P}(\xi_{T+h} = x \mid \boldsymbol{\xi}_{1:T} = \mathbf{x}_{1:T}) &= \frac{\mathsf{P}(\xi_{T+h} = x, \boldsymbol{\xi}_{1:T} = \mathbf{x}_{1:T})}{\mathsf{P}(\boldsymbol{\xi}_{1:T} = \mathbf{x}_{1:T})} \\
&= \frac{\mathbf{uB}(x_1)\ldots\mathbf{AB}(x_T)\mathbf{A}^h\mathbf{B}(x)\mathbf{1}^\mathsf{T}}{\boldsymbol{\alpha}_T\mathbf{1}^\mathsf{T}} = \frac{\boldsymbol{\alpha}_T\mathbf{A}^h\mathbf{B}(x)\mathbf{1}^\mathsf{T}}{\boldsymbol{\alpha}_T\mathbf{1}^\mathsf{T}} \\
&= \bar{\boldsymbol{\alpha}}_T\mathbf{A}^h\mathbf{B}(x)\mathbf{1}^\mathsf{T}
\end{aligned}
$$

Также справедлив предел для эргодической МЦ

$$
\lim_{h\to\infty} \mathsf{P}(\xi_{T+h} = x \mid \boldsymbol{\xi}_{1:T} = \mathbf{x}_{1:T}) = \lim_{h\to\infty} \bar{\boldsymbol{\alpha}}_T\mathbf{A}^h\mathbf{B}(x)\mathbf{1} = \tilde{\mathbf{u}}\mathbf{B}(x)\mathbf{1}^\mathsf{T},
$$

где $\tilde{\mathbf{u}}$ — стационарное распределение МЦ.

Детали реализации

- R reference classes.
- Не зависящие от конкретных плотностей алгоритмы, класс `Hmm`.
- Дочерние классы:
  - `PoisHmm`: СММ с пуассоновскими плотностями компонент смеси;
  - `NormHmm`: СММ с нормальными плотностями компонент смеси;
  - `MNormHmm`: СММ с многомерными нормальным плотностями компонент смеси;
  - `CategHmm`: «классическая» СММ для анализа категориальных данных, задаваемыми дискретными распределениями (попросту, таблицей).

Оценка параметров вырожденного случая

1. Смоделировать выборку из `PoisHmm(2)` с параметрами:

$$\mathbf{A}_{\text{ref}} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{u}_{\text{ref}} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \boldsymbol{\lambda}_{\text{ref}} = \begin{pmatrix} 6 & 1 \end{pmatrix}.$$

2. Инициализировать другую `PoisHmm(2)`.
3. Оценить параметры по ММП. После 38 итераций:

$$\hat{\mathbf{A}} = \begin{pmatrix} 0.98 & 0.02 \\ 1 & 0 \end{pmatrix}, \quad \hat{\mathbf{u}} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad \hat{\boldsymbol{\lambda}} = \begin{pmatrix} 6.06 & 0 \end{pmatrix}.$$

- Корпус российских новостей за 2008 год привести к виду

$$\mathbf{x}_{1:80} = \text{"без единой тенденции закрылись торги"}$$

- $\texttt{CategHmm}(2)$, начальные параметры:

$$\mathbf{A} = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} 0.03 & \dots & 0.03 \\ 0.03 & \dots & 0.03 \end{pmatrix}, \ \mathbf{u}^{\mathsf{T}} = \begin{pmatrix} 0.48 \\ 0.52 \end{pmatrix}.$$

- Оценить параметры по ММП на $\mathbf{x}_{1:1000}$.

Категориальная СММ на корпусе российских новостей

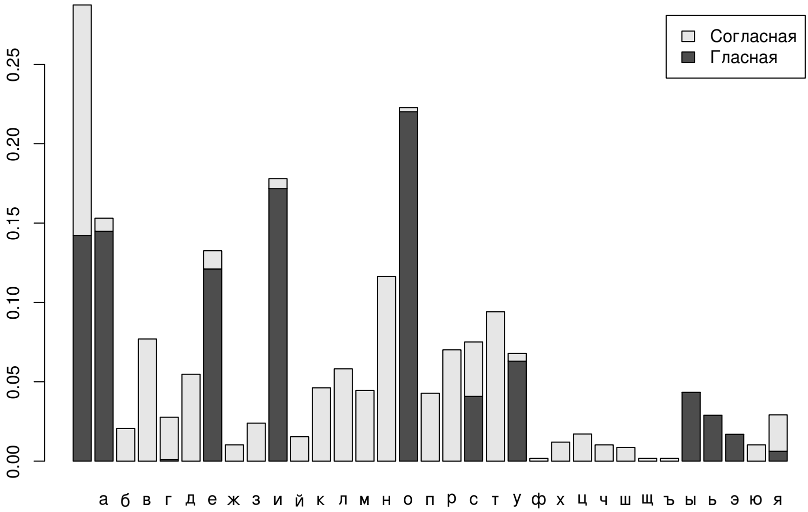


Рис.: Матрица $\mathbf{B}$ (успешная оценка)

Категориальная СММ на корпусе российских новостей

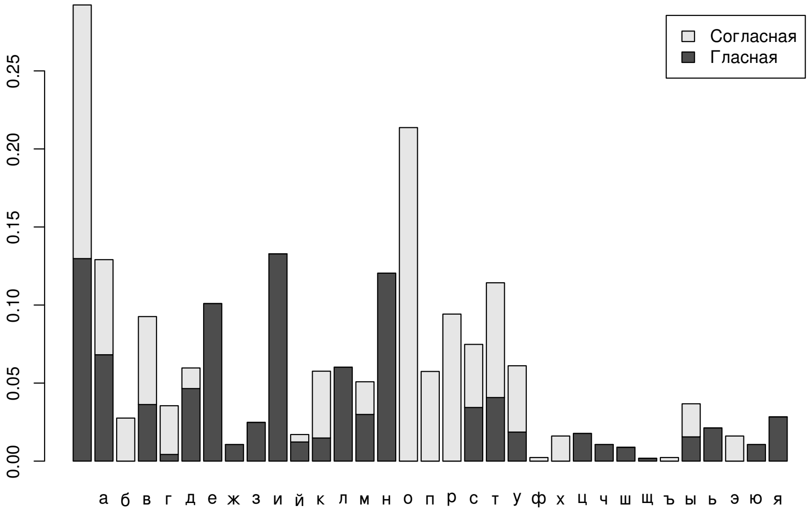
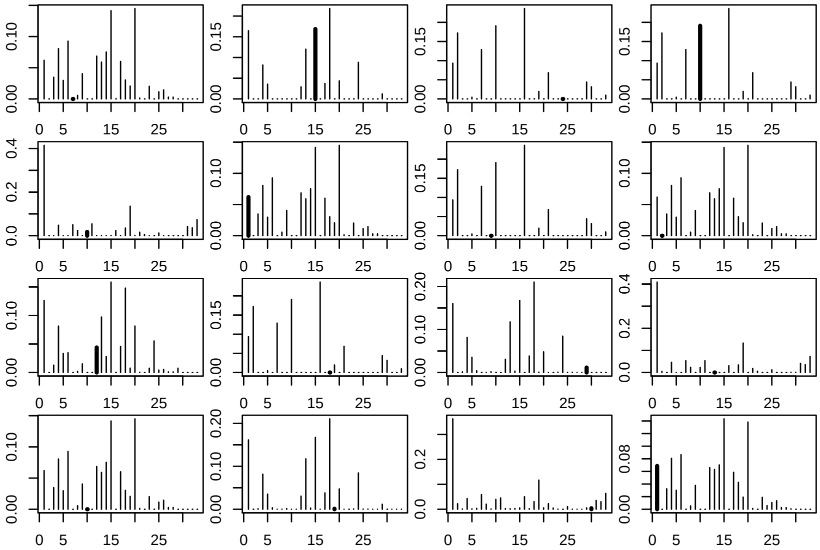


Рис.: Матрица $\mathbf{B}$ (неуспешная оценка)

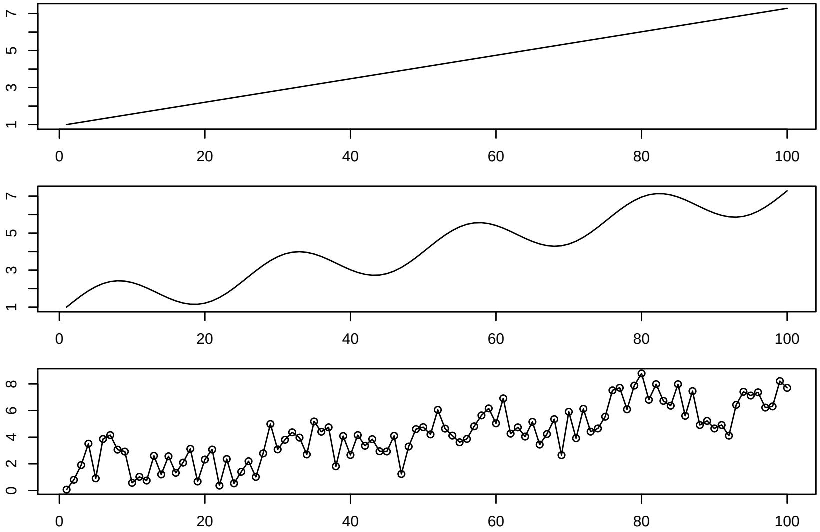
- Если элементы на *побочной* диагонали больше вне-диагональных, то алгоритм корректно оценивает $\mathbf{B}$. Пример:
$$\mathbf{A}_{\text{init}} = \begin{pmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{pmatrix}.$$
- Если элементы на *главной* диагонали больше вне-диагональных, то алгоритм застревает в локальном минимуме. Пример:
$$\mathbf{A}_{\text{init}} = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}.$$

Интерполяция на корпусе новостей



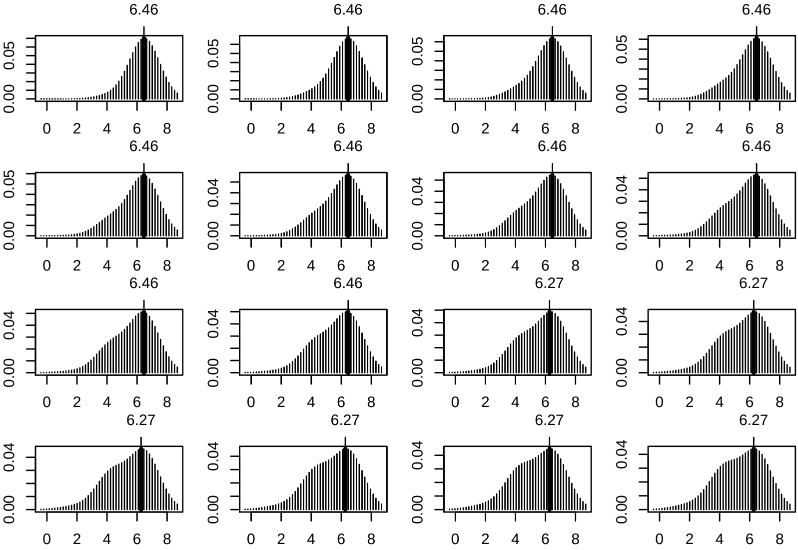
% угадано хотя бы по одной характеристике: 11.9
% угадано по моде: 7.6
% угадано по среднему: 1.8
% угадано по медиане: 2.5

Синтетические данные
Экстраполяция



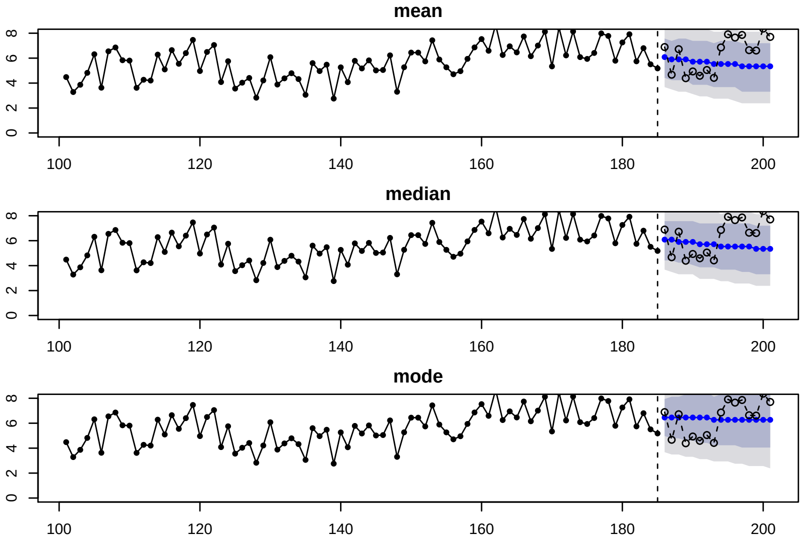
Распределение прогноза NormHMM(4)

Экстраполяция



Прогноз NormHMM(4) с тестовыми данными

Экстраполяция



Прогноз курса акций Apple
По (Nguen, 2016)

Пусть даны дневные цены акций AAPL:

open_t цена открытия в день t

high_t максимальная цена за день t

low_t минимальная цена за день t

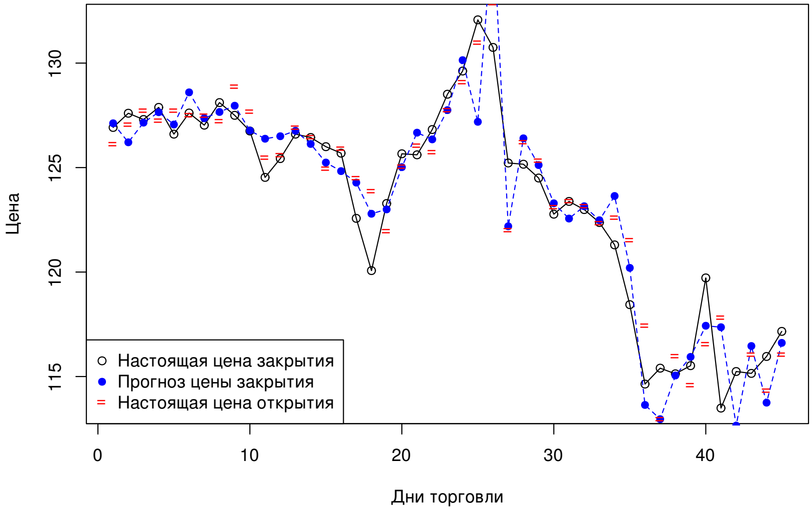
close_t цена закрытия в день t

Ряд $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, где $\mathbf{x}_t = (\mathrm{open}, \mathrm{high}, \mathrm{low}, \mathrm{close})$.

Алгоритм оценки close_{T+1}

1. Найти $\hat{\mathcal{M}}_T$ СММ `MNormHmm(2)` .
2. Найти такое t^*, что $\mathsf{P}(\mathbf{X}_{T-\ell:T}) \approx \mathsf{P}(\mathbf{X}_{t^*-\ell:t^*})$, ℓ — длина окна.
3. Найти

$$\begin{aligned}\widehat{\mathrm{close}}_{T+1} &= \mathrm{close}_T + (\mathrm{close}_{t^*+1} - \mathrm{close}_{t^*}) \\ &\quad \times \mathrm{sign}(\mathsf{P}(\mathbf{X}_{t^*-\ell:t^*}) - \mathsf{P}(\mathbf{X}_{T-\ell:T})).\end{aligned}$$



130
125
120
115
Цена
0
10
20
30
40
Дни торговли
Настоящая цена закрытия
Прогноз цены закрытия
Настоящая цена открытия

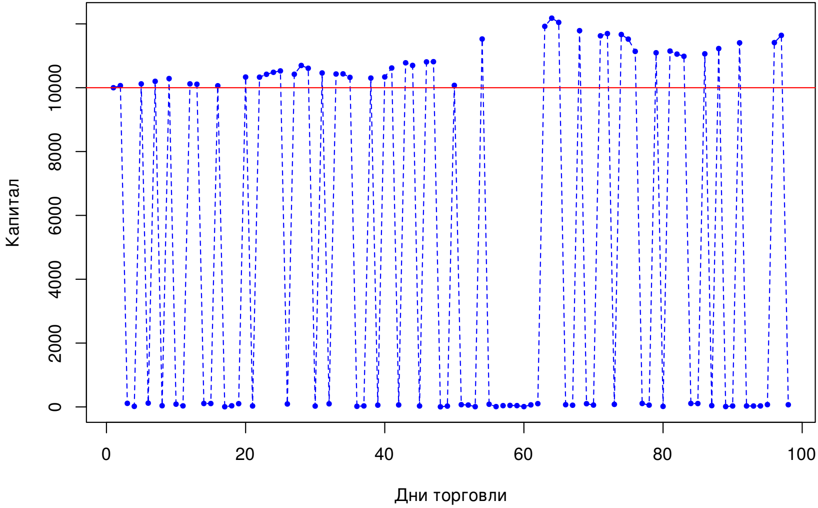
Проверка на простейшей торговой стратегии

Пусть P — прогноз цены закрытия, O — настоящая цена открытия.

Алгоритм

1. По известной O предсказать P;
2. если $O > P$ (акции падают в цене), то в начале дня следует продать акции, а в конце дня купить;
3. если $O < P$, то наоборот.

Проверка на простейшей торговой стратегии



Заключение

- Изучен формализм СММ как зависимой смеси распределений, в частности, способы
  - матричного представление алгоритмов;
  - задания произвольных плотностей для процесса наблюдений;
  - численной максимизация функции правдоподобия;
  - рассмотрения функционала, связанного с выводом и предсказанием наблюдений;
  - оптимального выбор компонент смеси через информационные критерии.
- Создана объектно-ориентированной реализации на R.
- Протестирован алгоритм оценки параметров на модельных данных для разных распределений компонент смеси.
- Изучена работоспособность алгоритмов интер- и экстраполяции.

Если $\operatorname{dom}\theta \neq \mathbb{R}$, найти обратимое

$$\begin{aligned} \phi : \operatorname{dom}\theta &\to \mathbb{R} \\ \theta &\mapsto \tilde{\theta}. \end{aligned}$$

Примеры

- $\theta_i > 0$:
$$\phi = \log : (0, \infty) \to \mathbb{R}$$
- $\mathbf{A}\mathbf{1}^{\mathsf{T}} = \mathbf{1}^{\mathsf{T}},\ a_{ij} \geq 0\ \forall i, j \in 1:m$:
$$\phi^{-1} : \mathbb{R} \to \mathbb{R}^{+}, \quad \text{строго монотонная.}$$
- $\mathbf{\Sigma} = (\sigma_{ij})_{i,j=1}^{d}$ — симметричная, неотр. опр.: разл. Холецкого $\mathbf{\Sigma} = \mathbf{L}^{\mathsf{T}}\mathbf{L}$ + логарифмирование (Pinheiro,1996):
$$\tilde{\mathbf{\Sigma}} = (\log\sigma_{11}, \sigma_{12}, \log\sigma_{22}, \sigma_{13}, \sigma_{23}, \log\sigma_{33}, \dots, \log\sigma_{dd}).$$

Так как $\sum_{j=1}^m a_{ij} = 1,\ \forall i \in 1:m,\ a_{ii} = 1 - \sum_{j\neq i} a_{ij}$, имеем

$$\begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mm} \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & \dots & \frac{a_{1m}}{a_{11}} \\ \vdots & \ddots & \vdots \\ \frac{a_{m1}}{a_{mm}} & \dots & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} \tilde{a}_{11} & \dots & \phi(\frac{a_{1m}}{a_{11}}) \\ \vdots & \ddots & \vdots \\ \phi(\frac{a_{m1}}{a_{mm}}) & \dots & \tilde{a}_{mm} \end{pmatrix}.$$

Обратно, $a_{ij} = \phi^{-1}(\tilde{a}_{ij})a_{ii}$, и a_{ii} находится из условия

$$\sum_{j=1}^m \phi^{-1}(\tilde{a}_{ij})a_{ii} = 1,$$

откуда

$$a_{ii} = \frac{1}{\sum_{j=1}^m \phi^{-1}(\tilde{a}_{ij})}, \quad a_{ij} = \frac{\phi^{-1}(\tilde{a}_{ij})}{\sum_{j=1}^m \phi^{-1}(\tilde{a}_{ij})}.$$

Состятельность ОМП

Утверждение

Если $\mathcal{M}^*$ — истинное значение параметров, $\hat{\mathcal{M}}_T$ — ОМП оценка по выборке объема T, то, *при некоторых условиях*,

$$\hat{\mathcal{M}}_T \xrightarrow[T\to\infty]{} \mathcal{M}^* \quad \text{п.в.}$$

$\operatorname{supp} \kappa_t$	$\operatorname{supp} \xi_t$	Авторы
Конечный	Конечный	(Baum et al., 1966)
Конечный	Произвольный	(Leroux, 1992)
Польское	Польское	(Douc et al., 2011)

Алгоритм оценки параметров чувствителен к начальным значениям.

Инициализация параметров

$\mathbf{A}$ равномерно с зашумлением либо так, что $a_{ii} = r a_{ij}$

$\boldsymbol{\theta}_j$ кластеризовать выборку; в каждом из m кластеров посчитать соответствующую выборочную характеристику

Инициализация $\boldsymbol{\theta}_j$

Пусть $\xi_t \sim \mathrm{Pois}(\lambda_j)$. Поскольку $\lambda = \mathsf{E}\xi$, то $\hat{\lambda}_j = \bar{\mathbf{x}}^{(j)}$, где $\mathbf{x}^{(j)}$ — j-й кластер.

Пусть даны дневные цены акций AAPL:

open_t цена открытия в день t

high_t максимальная цена за день t

low_t минимальная цена за день t

close_t цена закрытия в день t

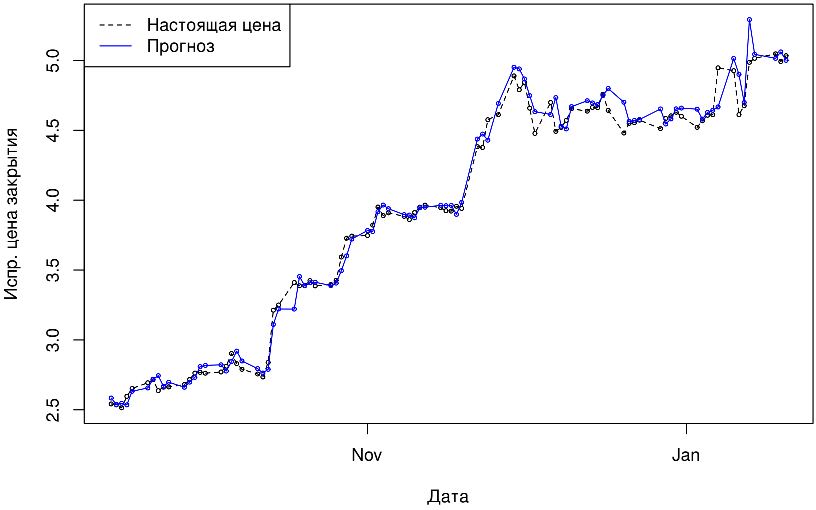
Алгоритм оценки close_{T+1}

1. Составить ряд $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_T\}$, где

$$\mathbf{x}_t = \left(\frac{\mathrm{close}_t - \mathrm{open}_t}{\mathrm{open}_t}, \frac{\mathrm{high}_t - \mathrm{open}_t}{\mathrm{open}_t}, \frac{\mathrm{open}_t - \mathrm{low}_t}{\mathrm{open}_t} \right).$$

2. Найти $\hat{\mathcal{M}}_T$.
3. Найти $\mathbf{x}^*_{T+1} = \operatorname{argmax}_{\mathbf{x}} \mathsf{P}(\xi_{T+1} = \mathbf{x} \mid \boldsymbol{\xi}_{(T-\ell):T} = \mathbf{x}_{(T-\ell):T})$.
4. Найти $\widehat{\mathrm{close}}_{T+1} = \left[x^*_{T+1}\right]_1 \cdot \mathrm{open}_{T+1} + \mathrm{open}_{T+1}$.

Прогноз курса акций Apple
По (Gupta, 2012)



Прогноз курса акций Apple
По (Gupta, 2012)