UNIVERSITY OF TECHNOLOGY
IN THE EUROPEAN CAPITAL OF CULTURE
CHEMNITZ

Faculty of Mathematics
Mathematical Finance
Alois Pichler

Forecasting Future Values From Time Series Data

Perspective: *Seminar thesis, Bachelor, Master's thesis*
Writing hints: https://www.overleaf.com/read/hdpkgxgjkbgw

Preliminaries

Consider a random variable (random vector) $X \in \mathbb{R}^d$ with density f_X,

$$f_X(x)\,\mathrm{d}x = P(X \in \mathrm{d}x), \qquad x \in \mathbb{R}^d.$$

The *Rosenblatt–Parzen* kernel density estimator for the density function f_X is

$$\hat{f}_n(x) := \frac{1}{n}\sum_{i=1}^n k_h(x - X_i), \qquad x \in \mathbb{R}^d,$$

where X_i, $i = 1, \dots, n$, are independent observations of X. The kernel $k\colon \mathbb{R}^d \to \mathbb{R}$ is non-negative ($k(x) \geq 0$ for all $x \in \mathbb{R}^d$) and satisfies $\int_{\mathbb{R}^d} k(x_1, \dots, x_d)\,\mathrm{d}x_1 \dots \mathrm{d}x_d = 1$.[1] For a strictly positive definite bandwidth matrix H, the kernel with bandwidth is

$$k_H(x) := \frac{1}{|\det H|} k\big(H^{-1}x\big).$$

Most often, $H = \begin{pmatrix} h_1 & 0 & \ddots \\ 0 & \ddots & 0 \\ \ddots & 0 & h_d \end{pmatrix}$ with the scalar $h := h_1 = \dots = h_d$ termed the *bandwidth*. A value of the bandwidth h too close to zero leads to *overfitting*, larger values for λ lead to *oversmoothing* (underfitting).

The conditional density is

$$f(x|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad x \in \mathbb{R}^d, y \in \mathbb{R}^\ell, \tag{1}$$

where $f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y = P(X \in \mathrm{d}x, Y \in \mathrm{d}y)$ is the joint density of the random vector $(X,Y) \in \mathbb{R}^d \times \mathbb{R}^\ell$ (which we assume to exist here for convenience).

[1] Popular kernels include the sigmoid kernel $k(x) := \frac{2}{\pi}\frac{1}{e^{\|x\|}+e^{-\|x\|}}$, which nowhere vanishes and thus is often better suited than the Epanechnikov kernel in the present context.

Employing two kernels $k\colon \mathbb{R}^d \to \mathbb{R}$ and $k\colon \mathbb{R}^\ell \to \mathbb{R}$, the estimator for the conditional density (1) is

$$\hat{f}_n(x|\, y) \coloneqq \sum_{i=1}^n k_h(x - X_i) \cdot \underbrace{\frac{k_h(y - Y_i)}{\sum_{j=1}^n k_h(y - Y_j)}}_{=: w_i(y)}, \tag{2}$$

where (X_i, Y_i), $i = 1, \dots, n$, are independent observations of the pair (X, Y); the weights $w_i(y)$ are non-negative and $\sum_{i=1}^n w_i(y) = 1$.

Samples from the conditional distribution

To obtain random samples from the conditional distribution $f(\cdot|\, y)$ in (2) for known y recall that $\sum_{i=1}^n w_i(y) = 1$. We employ the *composition method*: for a uniform $U \in [0, 1]$, find the index $i^* \in \{1, \dots, n\}$ such that

$$w_1(y) + \dots + w_{i^*-1}(y) < U \le w_1(y) + \dots + w_{i^*}(y),$$

and note that the random index i^* satisfies $P(i^* = i) = w_i(y)$, $i = 1, \dots, n$. Then

$$X_{i^*} + h \cdot \varepsilon \sim \hat{f}_n(\cdot|\, y), \tag{3}$$

that is, $X_{i^*} + h \cdot \varepsilon$ is random with density $\hat{f}_n(\cdot|\, y)$, provided that ε is chosen randomly from the distribution with density k, the random index i^* is as above, and $h > 0$ is the scalar bandwidth.

Application to Non-Linear Time Series

Suppose a stationary time series (X_t), $X_t \in \mathbb{R}$, with lag ℓ has conditional density

$$f(x_{\ell+1}|\, x_1, \dots, x_\ell). \tag{4}$$

Based on observing the trajectory

$$X_{2-\ell}, \dots, X_{n+1} \tag{5}$$

of the time series, the estimator for the conditional density (4) is

$$\hat{f}_n(x_{\ell+1}|\, x_1, \dots, x_\ell) = \sum_{i=1}^n k_h(x_{\ell+1} - X_{i+1}) \cdot \underbrace{\frac{k_h(x_1 - X_{i+1-\ell}) \cdot \ldots \cdot k_h(x_\ell - X_i)}{\sum_{j=1}^n k_h(x_1 - X_{j+1-\ell}) \cdot \ldots \cdot k_h(x_\ell - X_j)}}_{w_i(x_1, \dots, x_\ell)}. \tag{6}$$

UNIVERSITY OF TECHNOLOGY
IN THE EUROPEAN CAPITAL OF CULTURE
CHEMNITZ

Faculty of Mathematics
Mathematical Finance
Alois Pichler

Task: Forecasting with Delayed Embedding of Dimension ℓ

Suppose observations $X_1, \dots, X_\ell$ are given. We may sample $X_{\ell+1}$, the next observation of the time series, by employing (3).

1. Discuss and implement the relations (1)–(6) for a suitable, oscillating time series with (X_t) (for example the *nottem* time series data, which is included in the statistical programming language *R*).

2. **Recursive multistep forecasting and sliding window**: For some chosen, initial starting values $(\hat{X}_1, \dots, \hat{X}_\ell)$, extend this time series by getting new samples from (6) with (3) and *plot* the realization $\hat{X}_1, \hat{X}_2, \dots, \hat{X}_\ell, \dots, \hat{X}_n, \dots, \hat{X}_N$ with some $N \gg \ell$.

3. Adjust the kernel and the scalar bandwidth h so that the time series $\hat{X}$ *visually* matches the initial observations (5) of X.

Bonne chance!