Nonparametric quantile estimation for time series data

Dana Draghicescu
WSL, Landscape Department
Zürcherstrasse 111
8903 Birmensdorf Switzerland
dana.draghicescu@wsl.ch

Sucharita Ghosh
WSL, Landscape Department
Zürcherstrasse 111
8903 Birmensdorf Switzerland
rita.ghosh@wsl.ch

1. Motivation

This research was motivated by the differing patterns observed in the probability distribution functions of several climatic variables in Switzerland. An extensive exploratory data analysis performed on records of daily precipitation (in mm) during 1901-1999 at 114 locations in Switzerland showed that the distribution functions may have experienced temporal changes, while analysis of the data from individual stations reveal that the patterns need not be the same. As an illustration, Figure 1 displays the time series of yearly winter maxima in three locations that are representative for their geographic regions: Zürich (North), Davos (Alpine) and Locarno (South) respectively (top), together with their autocovariance functions (bottom).

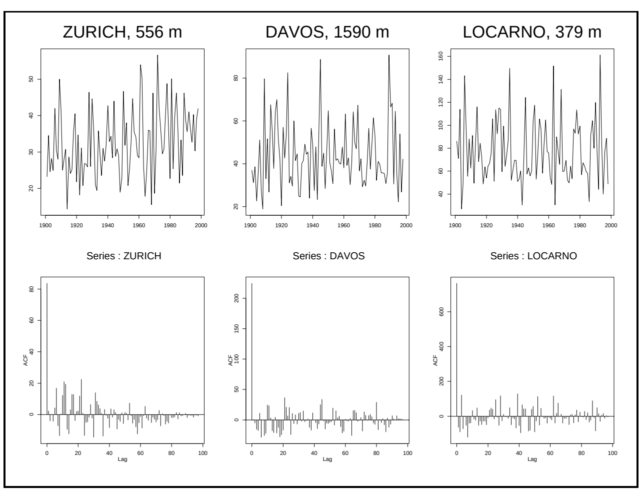

Figure 1: Yearly winter maximum precipitation (mm) during 1901-1999

2. Estimation

Suppose we observe a process of the form $Y_i = G(Z_i, x_i),\ i = 1, 2, \ldots,$ where $\{Z_i\}$ is a stationary zero mean Gaussian process and G is a smooth function. The time dependent probability distribution function $F_x(y) = P(Y(x) \leq y)$ can be estimated nonparametrically by $\hat{F}_x(y) = \frac{1}{nb}\sum_{i=1}^{n} K\left(\frac{x_i - x}{b}\right) 1_{\{Y_i \leq y\}}$ (see Ghosh et al. (1997)), where it is assumed that $\frac{\partial}{\partial x}F_x(y) = f_x(y)$ and $\frac{\partial^2}{\partial x^2}F_x(y)$ exist, K is a symmetric density function with support $[-1, 1]$ for which second derivatives exist, the bandwidth $b = b(n)$ is such that, as $n \to \infty$, $b \to 0$ and $nb \to \infty$ and $x_i = \frac{i}{n}$ are rescaled time points. Some first results were presented in Draghicescu & Ghosh (2000). The same estimator was used for predicting distribution functions for long-memory processes in Ghosh & Draghicescu (2001).

In this paper we introduce a locally optimal bandwidth selection algorithm. Essentially it uses estimates of the time dependent bias and variance of $\hat{F}_x(y)$ computed from the data. The integrated mean square error of $\hat{F}_x(y)$ can be written as $IMSE(x) \approx \int_{\mathbf{R}} \left[A^2(x, y)b_x^4 + \frac{B_n(x, y; b_x)}{n^2 b_x^2}\right] dy$, where A and B_n stem from the bias and variance of $\hat{F}_x(y)$ respectively. The iterative procedure for optimal bandwidth selection is based on this $IMSE(x)$. The optimal bandwidths thus obtained are then used to construct estimated quantile curves as $\hat{\theta}_\xi(x) = inf\{y : \hat{F}_x(y) \geq \xi\}$ for $0 < \xi < 1$ and $x \in [0, 1]$.

REFERENCES

Draghicescu, D. and Ghosh, S. (2000). Different aspects of quantile estimation for transformations of Gaussian processes. Abstracts. *Annual Symposium on Forecasting, Lisbon, June 2000.*

Ghosh, S., Beran, J. and Innes, J. (1997). Nonparametric conditional quantile estimation in the presence of long memory. *Student*, **2**, 109-117.

Ghosh, S. and Draghicescu, D. (2001). Predicting the distribution function for long-memory processes. To appear in the *International Journal of Forecasting.*

RESUME

La fonction de répartition marginale d'un processus stochastique - fonction dépendant du temps - est estimée de manière non paramétrique en utilisant lissage par noyaux. Dans ce travail, nous considérons un procédé de sélection de paramètres de lissage optimaux dirigés par les données. Ce procédé permet d'obtenir des estimateurs de courbes quantiles. Les résultats sont illustrés par des applications à quelques relevés de précipitations en Suisse.