**On the thick-pen transformation for time series**

Piotr Fryzlewicz

(joint work with Hee-Seok Oh)

Traditional visualisation of time series data consists of plotting the time series values against time and "connecting the dots". We propose an alternative, multiscale visualisation technique, motivated by the scale-space approach in computer vision. In brief, our method also "connects the dots", but uses a range of pens of varying thicknesses for this purpose. The resulting multiscale map, termed the Thick-Pen Transform (TPT) corresponds to viewing the time series from a range of distances. We hope that the resulting set of plots will provide interesting and useful information about the structure of the time series, not only in a heuristic, but also in a formal probabilistic sense.

We define the TPT of a real-valued univariate process $(X_t)_{t=1}^n$ as follows. Let $\mathcal{T}$ denote the set of thickness parameters. For each $\tau_i \in \mathcal{T}$, $i = 1, \ldots, |\mathcal{T}|$, let $U_t^{\tau_i}$ denote the upper boundary of the area covered by a pen of thickness $\tau_i$ while connecting the points $(t, X_t)_{t=1}^n$. Similarly, let $L_t^{\tau_i}$ denote its lower boundary. The TPT $TP_{\mathcal{T}}(X_t)$ is the sequence of all pairs of boundaries, i.e.

$$TP_{\mathcal{T}}(X_t) = \{(L_t^{\tau_i}, U_t^{\tau_i})_{t=1}^n\}_{i=1,\ldots,|\mathcal{T}|}.$$

The precise mathematical form of $TP_{\mathcal{T}}(X_t)$ depends on the shape of the pen used. For example, consider a pen which is a closed square of side length $\tau \in \mathcal{T}$, positioned so that two of its sides are parallel to the time axis. For each point along the straight line connecting $(t, X_t)$ with $(t+1, X_{t+1})$, we place the pen so that the given point is at the centre of the right-hand side of the pen. In this set-up, we have

$$U_t^\tau = \max(X_t, \ldots, X_{t+\tau}) + \frac{\tau}{2} \quad (1)$$

$$L_t^\tau = \min(X_t, \ldots, X_{t+\tau}) - \frac{\tau}{2}. \quad (2)$$

Other pen shapes are possible, in the same way that a variety of kernel shapes are possible in kernel smoothing. Considering $TP_{\mathcal{T}}(X_t)$ for a range of thickness values $\tau$, a multiscale transform of the data $X_t$ is obtained, with higher values of the thickness parameter bringing out coarser-scale features of the data, and vice versa.

The TPT is not the only multiscale tool in time series analysis. Wavelets, which provide linear, multiscale and local decomposition of data, have been used extensively in time series analysis (see e.g. [5]). SiZer ([1]) is a linear data visualisation technique for displaying features of kernel-smoothed data as a function of location and bandwidth, simultaneously over a range of bandwidths. We note here that the TPT is not linear as it is based, effectively, on localised and weighted min/max operations. Self-similarity and (multi-)fractality are oft-recurring concepts in time series analysis, aiming to study parametric relationships between distributions of the process at different scales, particularly in the context of long-range dependent processes, see e.g. [2]. Besides using different methodology, the aims of the TPT

are different: we regard it as a visualiser which can be applied to any time series and which can ultimately assist in solving tasks such as nonstationarity detection, classification or measuring dependence between time series.

In the TPT as described above, one thickness value $\tau$ generates two sequences: $U_t^\tau$ and $L_t^\tau$. In some time series problems, it might be more convenient to use a single summary sequence, instead of a pair. Probably the simplest possible summary sequences involving $U_t^\tau$ and $L_t^\tau$ are

- Volume of the pen, defined as $V_t^\tau = U_t^\tau - L_t^\tau$;
- Mean of the pen, defined as $M_t^\tau = \frac{1}{2}\{U_t^\tau + L_t^\tau\}$.

Many more summary statistics are possible, also those combining $U_t^\tau$ and $L_t^\tau$ nonlinearly. The volume statistic $V_t^\tau$ deserves special attention as statistical literature has previously explored the concept of "the volume of a covering of data", albeit in other contexts. [7] derived the "tube formula" for calculating the volume of a tube surrounding a smooth manifold. This result has more recently been applied in various statistical contexts by a number of authors, see e.g. [6]. We are unaware of any applications of tube formulae in classical time series, where sample paths are often intrinsically non-smooth. On the other hand, in estimating the Hurst exponent or the fractal dimension of stochastic processes, two techniques involving statistics related to $V_t^\tau$ are the Rescaled Range Analysis ([4]) and the "box-counting" method, whose statistical properties in estimating the fractal dimension of a stationary continuous-time Gaussian process were studied in [3]. By contrast, our $V_t^\tau$ statistic is not an estimator, and applies to discrete-time, also nonstationary processes.

We now state a discrimination property of the TPT, which implies, roughly speaking, that two differently distributed Gaussian time series have differently distributed TPTs, under the (mild) Assumption 1 below. This is an important result as it gives us hope that the TPT can serve as an effective discriminant for time series.

**Assumption 1.** *For a given fixed lag $\tau > 0$, a process $X_t$ satisfies*

$$\exists\, \lambda_0, \delta \in [0,1) \quad \forall \lambda > \lambda_0 \quad \forall\, t$$
$$P\left( \bigcup_{t \le i,j \le t+\tau; \{i,j\} \neq \{t,t+\tau\}} |X_i - X_j| > |X_t - X_{t+\tau}| \;\middle|\; |X_t - X_{t+\tau}| > \lambda \right) \le \delta.$$

**Discrimination theorem.** *Let $X_t, Y_t$ be two zero-mean Gaussian time series such that for some $s < t$, the distribution of $X_s - X_t$ is not the same as the distribution of $Y_s - Y_t$, and let both $X_t$ and $Y_t$ satisfy Assumption 1 with $\tau = t - s$. Let $TP_{\mathcal{T}}(X_t)$, $TP_{\mathcal{T}}(Y_t)$ be the TPTs of $X_t$, $Y_t$ respectively, both with the square pen where the set $\mathcal{T}$ of thickness parameters is $\mathcal{T} = \{1, 2, \ldots\}$, and let $V_t^\tau(X), V_t^\tau(Y)$ be the corresponding volumes. Then, $TP_{\mathcal{T}}(X_t)$ and $TP_{\mathcal{T}}(Y_t)$ follow different probability distributions in the sense that the tri-variate random vectors*

$(V_s^{\tau-1}(X), V_{s+1}^{\tau-1}(X), V_s^{\tau}(X))$ *and* $(V_s^{\tau-1}(Y), V_{s+1}^{\tau-1}(Y), V_s^{\tau}(Y))$ *are distributed differently.*

We have applied the TPT to testing for time series stationarity, and to quantifying dependence between two time series. We introduce here the former application. The key result is as follows.

**Functional central limit theorem.** *Let* $\{X_t\}_{t=1}^n$ *be a stationary process satisfying* $\mathbb{E}|X_t|^r < \infty$ *for some* $r > 2$. *In addition let* $X_t$ *be* $\alpha$*-mixing with the mixing coefficients* $\alpha_m$ *satisfying* $\alpha_m = O(m^{-s})$ *for some* $s > \frac{r}{r-2}$. *Let* $TP_{\mathcal{T}}(X_t)$ *be the TPT of* $X_t$ *using an arbitrary pen but such that both* $U_t^{\tau}$ *and* $L_t^{\tau}$ *are functions of* $X_{t-C\tau}, \ldots, X_{t+C\tau}$ *only, for some* $C > 0$. *Further let the summary sequence* $K_t^{\tau}$ *be such that for each fixed* $\tau$, *we have* $n^{-1}\mathrm{Var}(\sum_{t=1}^n K_t^{\tau}) \to \sigma_{\tau}^2 < \infty$, *and* $|K_t^{\tau}| \leq A + B|\max(X_{t-C\tau}, \ldots, X_{t+C\tau})|$ *for some constants* $A, B > 0$, *possibly depending on* $\tau$. *Under these conditions, the following functional central limit result holds for each fixed* $\tau$. *Let* $u \in [0,1]$ *and denote* $Y_n^{\tau}(u) = \sigma_{\tau}^{-1} n^{-1/2} \sum_{t=1}^{\lceil nu \rceil} K_t^{\tau} - \mathbb{E}(K_t^{\tau})$. *We have*

$$Z_n^{\tau}(u) := Y_n^{\tau}(u) - \frac{\lceil nu \rceil}{n} Y_n^{\tau}(1) \xrightarrow{d} B_u^0,$$

*where* $B_u^0$ *is the standard Brownian bridge process on* $[0,1]$.

Our stationarity test is based on the fact that under the null hypothesis of stationarity, the range of the empirical version of $Z_n^{\tau}(u)$ is distributed as the range of Brownian bridge. As the test is derived from the TPT, which is a visualiser, it can be regarded as a "visual" one. Hence, it should come as no surprise that it operates under low moment assumptions, and is equally valid for linear and nonlinear processes. It appears to offer very good empirical performance.

## References


[1] P. Chaudhuri and J.S. Marron, *SiZer for exploration of structures in curves*, J. Am. Stat. Assoc. **94** (1999), 807-823.

[2] P. Doukhan, G. Oppenheim and M.S. Taqqu, editors, *Theory and Applications of Long-Range Dependence*, Birkhäuser (2003).

[3] P. Hall and A. Wood, *On the performance of box-counting estimators of fractal dimension*, Biometrika **80** (1993), 246-252.

[4] H.E. Hurst, *Long term storage capacity of reservoirs*, Trans. Am. Soc. Civil Engineers **116** (1951), 770-799.

[5] D.B. Percival and A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press (2000).

[6] J. Sun, *Tail probabilities of the maxima of Gaussian random fields*, Ann. Prob., **21** (1993), 34-71.

[7] H. Weyl, *On the volume of tubes*, Amer. J. Math., **61** (1939), 461-472.