Theorem 5.12 *Let the nonlinearity $\mu(u)$ be piecewise-Lipschitz. If the scales M and K of the estimate in (5.29) are selected using the formulas*

$$M = \tfrac{1}{3}\log_2 N \text{ and } K = 2M,$$

then the nonlinear order statistics algorithm converges to this nonlinearity globally, in the MISE error sense, with the best case rate

$$\operatorname{MISE}\hat{\mu}_{MK} = \mathcal{O}\left(N^{-2/3}\right),$$

or with the worst case rate

$$\operatorname{MISE}\hat{\mu}_{MK} = \mathcal{O}\left(N^{-1/3}\right).$$

These convergence rates are similar to those in the nonlinear QOS algorithm, *i.e.* they are independent of:

- the type of system dynamics,
- the correlation structure of the external noise (which can now have an unbounded distribution), and – in contrast to the QOS algorithms – independent of
- the smoothness of the probability density function of the input signal.

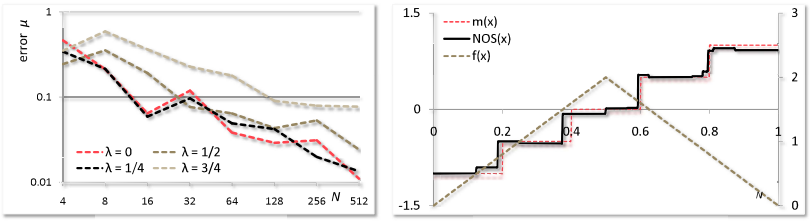

FIGURE 5.7. The *nonlinear order statistics* (NOS) estimate performance is strongly affected by the dynamics (left), but the estimate localizes the non-dyadic jumps well (right)

5.6 Empirical distribution (ED) algorithm

In the *empirical distribution algorithm* (ED), we recover $\mu(u)$ estimating an equivalent *composite function* $\mu_F \circ F(u) = \mu_F(F(u))$, where $F(u)$ is the *cumulative distribution function* of the input signal (having a probability density function $f(u)$).

Let $\{(u_k, y_k)\}$, $k = 1, \ldots, N$ denote the set of the system input-output measurements sorted pairwise w.r.t. to the increasing values of the inputs u_k. The algorithm is based on two observations:

- The *empirical distribution,* $F_N(u)$, generated by such ordered inputs u_k, maps the randomly scattered sequence $\{u_k\}$ onto the equidistant grid $\{x_k\}$:
$$x_k = F_N(u_k) = \tfrac{k}{N}. \tag{5.31}$$
- By virtue of Assumption 1, the cumulative distribution function $F(u)$ is continuous, strictly increasing (*viz.* invertible).

Throughout the chapter we assume, for simplicity, that N is a *dyadic integer* and hence $\{x_k\}$ forms a *dyadic grid* with grid points, $x_k = 2^{-\log_2 N} k$, being *binary rationals.*

The *empirical distribution estimate* of the nonlinearity $\mu(x)$ is of the form:
$$\hat{\mu}_K(x) = \sum_{n=0}^{2^K-1} \hat{\alpha}_{Kn} \varphi_{Kn}(x) \text{ where } x = F_N(u), \tag{5.32}$$
and where the empirical coefficients $\hat{\alpha}_{Kn}$ are the following estimates
$$\hat{\alpha}_{Kn} = \sum_{k=1}^{N} y_k \int_{\frac{k-1}{N}}^{\frac{k}{N}} \varphi_{Kn}(x)\, dx \tag{5.33}$$
of the unknown coefficients α_{Kn} of the nonlinearity $\mu(x)$ approximation in the space V_K; see Fig. 5.8 and *cf.* (4.10) and (5.24)):
$$\alpha_{Kn} = \int_0^1 \mu(x)\, \varphi_{Kn}(x)\, dx = \sum_{k=1}^{N} \int_{\frac{k-1}{N}}^{\frac{k}{N}} \mu(x)\, \varphi_{Kn}(x)\, dx. \tag{5.34}$$

The algorithm based on similar application of the empirical distribution function has been proposed in the statistical literature; see [25]. Moreover, the empirical coefficients used there are calculated using the formula
$$\check{\alpha}_{Kn} = \frac{1}{N} \sum_{k=1}^{N} y_k \varphi_{Kn}(x_k), \tag{5.35}$$
which seems to be simpler to compute than ours. Nevertheless, applying (like in OS algorithm) the *fundamental theorem of calculus* (see *e.g.* [1, Th. 5.3]), we have that
$$\int_{\frac{k-1}{N}}^{\frac{k}{N}} \varphi_{Kn}(x)\, dx = \Phi_{Kn}\left(\tfrac{k}{N}\right) - \Phi_{Kn}\left(\tfrac{k-1}{N}\right),$$
where $\Phi_{Kn}(x)$ are *indefinite integrals* of the scaling functions, $\varphi_{Kn}(x)$, and the resulting equivalent form of our estimate:
$$\hat{\alpha}_{Kn} = \sum_{k=1}^{N} y_k \left[\Phi_{Kn}\left(\tfrac{k}{N}\right) - \Phi_{Kn}\left(\tfrac{k-1}{N}\right)\right], \tag{5.36}$$

is, in fact, equally simple since, in case of the Haar function, the indefinite integrals, $\Phi_{Kn}(x)$ have the common compact form $\Phi_{Kn}(x) = \sqrt{2^{-K}}\Phi\left(2^K x - n\right)$, where $\Phi(x) = x \cdot \chi_{[0,1)}(x) + \chi_{[1,\infty)}(x)$; *cf.* (5.25). Observing further that

$$\check{\alpha}_{Kn} = \sum_{k=1}^{N} y_k \int_{\frac{k-1}{N}}^{\frac{k}{N}} \varphi_{Kn}(x_k)\, dx,$$

we can expect in general a better performance of our estimate due to its more accurate approximation of the integral in (5.34); *cf.* Remark 5.6 in the previous chapter.[11]

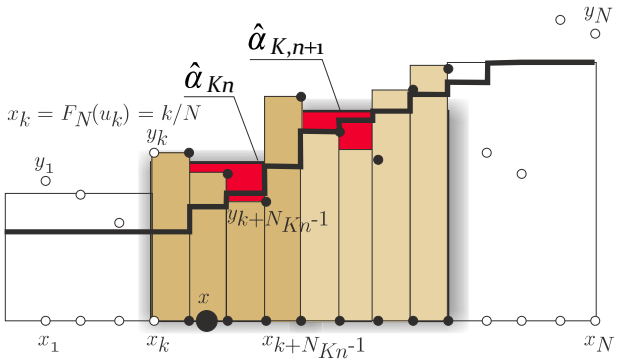

FIGURE 5.8. The illustration of the idea behind the ED algorithm. Note that input measurements are equidistant and each empirical coefficient is computed using the same number of measurements (weighted by the same factor $1/N$)

Note that the number of measurements used to evaluate each of the ED algorithm empirical coefficient $\hat{\alpha}_{Kn}$ equals $2^{-K}N$ and thus is fixed, deterministic and the same for all coefficients; *cf.* (A.16). In effect, the algorithm scale adapts to the local density of the input measurements (the local scale grows in regions where the number of measurement is larger than in the others; *cf.* Figs. 5.8, 5.9 and 7.1). This is a significant property which differentiates the ED algorithm from *e.g.* the the previous ones, where the number of measurements used to compute each of the empirical coefficients is random and changes from coefficient to coefficient; *cf.* (5.9) and (5.23).

5.6.1 Convergence

The mapping made by the empirical distribution $F_N(u)$ turns the randomly scattered inputs into the equidistant grid and may suggest that the initially random setting design regression estimation problem is turned into the fixed

[11] For dyadic N these formulas are actually equivalent - see Chapter 6.1.3.