UNIVERSAL PIECEWISE LINEAR REGRESSION OF INDIVIDUAL SEQUENCES: LOWER BOUND

Suleyman S. Kozat

IBM, Yorktown Heights, NY

Andrew C. Singer, Georg C. Zeitler

Department of ECE, University of Illinois, Urbana, IL

ABSTRACT

We consider universal piecewise linear regression of real valued bounded sequences under the squared loss function. In this setting, we present a lower bound on the regret of a universal sequential piecewise linear regressor compared to the best piecewise linear regressor that has access to the entire sequence in advance. This lower bound are tight with the corresponding upper bounds, suggesting a min-max optimality of the sequential regressor, for every individual bounded sequence.

Index Terms— Regression, piecewise linear, universal

1. INTRODUCTION

Consider the problem of piecewise linear p-th order regression of an arbitrary real-valued sequence. Both the outcome sequence x^n and the observation sequence y^n are assumed to be deterministic individual sequences which are bounded such that $|x[t]| < A_x$ and $|y[t]| < A_y$ for all t. At each time instant t, after forming an estimate $\hat{x}[t]$ based on observations $y[t-p+1], y[t-p], \ldots, y[t]$, one observes the t-th sample $x[t]$ of the sequence x^n. The accumulated loss of the regressor $\hat{x}[t]$ with respect to the sequence x^t up to time t is given by $l(x^t, \hat{x}^t) = \sum_{k=1}^{t}(x[k]-\hat{x}[k])^2$. This regressor is strongly sequential in the sense that at time t, it has only access to the observations $y[1], y[2], \ldots, y[t]$ up to time t and the past values of the outcome sequence, i.e., $x[1], x[2], \ldots, x[t-1]$. The goal of the sequential regressor is to perform almost as well as the best batch regressor knowing the entire sequence x^n in advance.

Sequentially available regression and prediction algorithms as well as upper and lower bounds on the regret of those algorithms are, e.g., described in the machine learning literature [1, 2, 3], the signal processing literature [4, 5, 6] and the information theory literature [7].

2. PIECEWISE LINEAR REGRESSION AND UPPER BOUND ON THE REGRET

Restriction to linear regression algorithms considerably limits the modeling power of the regressor. Instead, we focus on piecewise linear regression, where we parse the past observation space $[-A_y, A_y]^p$ spanned by the observations into J regions $\mathcal{R}_j$, $j = 1, 2, \ldots, J$, whose boundaries are fixed and known, such that $\bigcup_{j=1}^{J} \mathcal{R}_j = [-A_y, A_y]^p$. Then, using a piecewise linear regression $\tilde{x}[t]$ from a sequential algorithm, we try to minimize

$$\sup_{x^n, y^n} \left\{ \sum_{t=1}^{n} (x[t] - \tilde{x}[t])^2 - \min_{\underline{w} \in \mathbb{R}^{Jp}} \sum_{t=1}^{n} (x[t] - \underline{w}_{s[t]}^{\mathrm{T}} \underline{y}[t])^2 \right\}, \tag{1}$$

where the state indicator variable $s[t] = j$ if $\underline{y}[t] = [y[t], y[t-1], \ldots, y[t-p+1]]^{\mathrm{T}} \in \mathcal{R}_j$. The vector $\underline{w} = [\underline{w}_1^{\mathrm{T}}, \underline{w}_2^{\mathrm{T}}, \ldots, \underline{w}_J^{\mathrm{T}}]^{\mathrm{T}}$ collects the J linear regression vectors $\underline{w}_j \in \mathcal{R}^p$.

In [6], a sequential piecewise linear regressor $\tilde{x}[t]$ is presented whose regret with respect to the best piecewise linear batch regressor of order p satisfies

$$\frac{1}{n}\sum_{t=1}^{n}(x[t]-\tilde{x}[t])^2 \leq \frac{1}{n} \min_{\underline{w}} \left\{ \sum_{t=1}^{n} (x[t] - \underline{w}_{s[t]}^{\mathrm{T}} \underline{y}[t])^2 + \delta \|\underline{w}\|_2^2 \right\} + \frac{pJA_x^2}{n} \ln\left(\frac{n}{J}\right) + O\left(\frac{1}{n}\right), \tag{2}$$

for any $x^n \in [-A_x, A_x]^n$, $y^n \in [-A_y, A_y]^n$ and $\delta > 0$. Defining J time vectors of length n_j, $t_j^{n_j} = \{t : s[t] = j\}$, and sequences $x_j^{n_j} = \{x[t_j[k]]\}_{k=1}^{n_j}$ and $\underline{y}_j^{n_j} = \{\underline{y}[t_j[k]]\}_{k=1}^{n_j}$, the regressor $\tilde{x}[t]$ achieving this bound is given by [6]

$$\tilde{x}[t] = \underline{\tilde{w}}_{s[t]}^{\mathrm{T}}[t-1]\underline{y}[t], \tag{3}$$

with

$$\underline{\tilde{w}}_j[t-1] = \left(\sum_{k=1}^{t} \underline{y}_j[k]\underline{y}_j^{\mathrm{T}}[k] + \delta_j I_p \right)^{-1} \sum_{k=1}^{t-1} \underline{y}_j[k] x_j[k], \tag{4}$$

where $\delta_j > 0$ is a positive constant and I_p the $p \times p$ identity matrix. As stated in Eq. (2), there exists a sequential piecewise linear regressor of order p whose regret is at most $O(n^{-1}\ln(n))$.

3. LOWER BOUND ON THE REGRET

In the following, we derive a lower bound for

$$\inf_{q\in\mathcal{Q}}\sup_{x^n,y^n}\left\{\sum_{t=1}^{n}(x[t]-\tilde{x}_q[t])^2-\inf_{\underline{w}}\sum_{t=1}^{n}(x[t]-\underline{w}_{s[t]}^{\mathrm{T}}\underline{y}[t])^2\right\}, \tag{5}$$

where $\mathcal{Q}$ is the class of all sequential regressors. This lower bound is tight with the upper bound given in Eq. (2), suggesting that the regressor described in Eq. (3) is optimal in a sense that no sequential regressor can do much better, in a min-max sense. This is stated in the following theorem.

Theorem: *Let x^n and y^n be individual bounded sequences with $|x[t]| < A_x$ and $|y[t]| < A_y$, and let $\tilde{x}_q[t]$ form the output of a sequential regression algorithm. Then*

$$\inf_{q\in\mathcal{Q}}\sup_{x^n,y^n}\frac{1}{n}\left\{\sum_{t=1}^{n}(x[t]-\tilde{x}_q[t])^2-\inf_{\underline{w}}\sum_{t=1}^{n}(x[t]-\underline{w}_{s[t]}^{T}\underline{y}[t])^2\right\} \geq \frac{pJA_x^2(1-\epsilon)}{n}\ln\left(\frac{n}{J}\right)-\frac{G}{n}-O\left(\frac{1}{n^2}\right), \tag{6}$$

where $\mathcal{Q}$ is the class of all sequential regressors, for all $0 < \epsilon \leq p/(3p-1)$ and a positive constant $G > 0$.

Hence, for every sequential regressor there exists a pair of sequences x^n and y^n such that the normalized accumulated regression error is at least $O(n^{-1}\ln(n))$ worse than that of the best batch regressor. The proof of the theorem is based on results in [3] for linear regression.

3.1. Proof of the Theorem

Defining $x_{\underline{w}}[t] = \underline{w}_{s[t]}^{\mathrm{T}}\underline{y}[t]$, we have for any distribution on x^n and y^n that

$$\inf_{q\in\mathcal{Q}}\sup_{x^n,y^n}\left\{l(x^n,\tilde{x}_q^n)-\inf_{\underline{w}}l(x^n,x_{\underline{w}}^n)\right\}\geq L(n), \tag{7}$$

where

$$L(n) := \inf_{q\in\mathcal{Q}}\mathrm{E}\left[l(x^n,\tilde{x}_q^n)\right]-\mathrm{E}\left[\inf_{\underline{w}}l(x^n,x_{\underline{w}}^n)\right]. \tag{8}$$

Hence, it is enough to lower bound $L(n)$ to find a lower bound on Eq. (5).We now consider the following distribution on x^n and y^n. The sequence y^n is constructed such that $s[t] = 1$ for the first n_1 points, $s[t] = 2$ for the next n_2 points up to the last n_J points, where $s[t] = J$. The constraint on n_j is that $\sum_{j=1}^{J} n_j = n$. Hence, we have J independent least squares problems in J regions, and we solve the computation of the lower bound for each region independently. In each region j, pick p random variables θ_{ji}, $i = 1, 2, \ldots, p$, from a beta distribution with parameters (C_j, C_j), given by

$$p(\theta_{ji}) = \frac{\Gamma(2\mathrm{C_j})}{\Gamma(C_j)\Gamma(C_j)}\theta_{ji}^{C_j-1}(1-\theta_{ji})^{C_j-1}. \tag{9}$$

At time instant m, $m \in \{1, 2, \ldots, n_j\}$, let the observation vector $\underline{y}[m]$ be $\underline{y}[m] = [0, 0, \ldots, 0, y_j, 0, \ldots, 0, 0]^{\mathrm{T}}$, where the only non-zero entry in $\underline{y}[m]$ is $y_j \neq 0$ at the i-th position with

$$i = \begin{cases} p & \text{if } m \bmod p = 0 \\ m \bmod p & \text{otherwise.} \end{cases} \tag{10}$$

Alternatively, the observation vector can be written as $\underline{y}[m] = y_j\underline{e}_i$, where $\underline{e}_i \in \mathbb{R}^p$ is the unit vector with the only one at the i-th position. Furthermore, the signal $x_j[i + (\ell-1)p]$ is $x_j[i+(\ell-1)p] = A_x$ with probability $(1-\theta_{ji})$ and $x_j[i+(\ell-1)p] = -A_x$ with probability θ_{ji}, for all $\ell \in \{1, 2, \ldots, \lceil(n_j+1-i)/p\rceil\}$, independently of the previous trials.

3.2. Loss of Sequential Regressor

First, we compute the expected loss of the best sequential regressor which under the squared error loss is given by the MMSE regressor

$$\begin{aligned}\hat{x}_{j,q}[m] &= \hat{x}_{j,q}\left[i+(\lceil m/p\rceil-1)p\right]\\ &= \mathrm{E}\left[x_j[i+(\lceil m/p\rceil-1)p]|x_{ji}^{\tilde{m}-1}\right]\\ &= \mathrm{E}\left[(1-2\theta_{ji})A_x|x_{ji}^{\tilde{m}-1}\right],\end{aligned} \tag{11}$$

where $\tilde{m}$ is defined as $\tilde{m} := \lceil m/p\rceil$. Here, the sequence $x_{ji}^{\tilde{m}-1} = \{x_j[i+(\ell-1)p]\}_{\ell=1}^{\tilde{m}-1}$, collecting the $\tilde{m}-1$ samples of x_j^{m-1} associated with i. Applying Bayes' rule to $p(\theta_{ji}|x_{ji}^{\tilde{m}-1})$ and using the properties of the beta distribution, we obtain that

$$\mathrm{E}\left[\theta_{ji}|x_{ji}^{\tilde{m}-1}\right] = \frac{\tilde{m}-1-N_a+C_j}{\tilde{m}-1+2C_j}, \tag{12}$$

with N_a denoting the number of occurrences of A_x in the sequence $x_{ji}^{\tilde{m}-1}$. Then, the best sequential regressor can be written as

$$\hat{x}_{j,q}[m] = \frac{\sum_{k=0}^{\tilde{m}-1}x_j[i+(k-1)p]}{\tilde{m}-1+2C_j}. \tag{13}$$

Defining n_{ji} as $n_{ji} := \lceil(n_j+1-i)/p\rceil$, the expected loss of this regressor in region j can then be computed as

$$\begin{aligned}&\sum_{m=1}^{n_j}\mathrm{E}\left[(x_j[m]-\hat{x}_{j,q}[m])^2\right] \\ &=\sum_{i=1}^{p}\sum_{\tilde{m}=1}^{n_{ji}}\mathrm{E}\left[(x_j[i+(\tilde{m}-1)p]-\hat{x}_{j,q}[i+(\tilde{m}-1)p])^2\right].\end{aligned} \tag{14}$$

Expanding the square, we find that

$$\mathrm{E}[x_{ji}^2(i+(\tilde{m}-1)p)] = \mathrm{E}[\mathrm{E}[x_{ji}^2[i+(\tilde{m}-1)p]|\theta_{ji}]] = A_x^2. \tag{15}$$

Since given θ_{ji}, $x_{ji}[i+(\tilde{m}-1)p]$ and $x_{ji}[i+(k-1)p]$, $1 \le k \le \tilde{m}-1$, are independent, we obtain that

$$\begin{aligned}
&\mathrm{E}\left[x_{ji}[i+(\tilde{m}-1)p]\sum_{k=1}^{\tilde{m}-1}x_{ji}[i+(k-1)p]\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[x_{ji}[i+(\tilde{m}-1)p]|\theta_{ji}\right]\mathrm{E}\left[\sum_{k=1}^{\tilde{m}-1}x_{ji}[i+(k-1)p]\middle|\theta_{ji}\right]\right] \\
&= A_x^2(\tilde{m}-1)\mathrm{E}\left[(1-2\theta_{ji})^2\right] \\
&= \frac{A_x^2(\tilde{m}-1)}{2C_j+1}. \qquad (16)
\end{aligned}$$

Finally, the second square term can be rewritten as

$$\begin{aligned}
&\mathrm{E}\left[\left(\sum_{k=1}^{\tilde{m}-1}x_{ji}[i+(k-1)p]\right)^2\right] \qquad (17)\\
&\qquad = \mathrm{E}\left[\mathrm{E}\left[\left(A_x(\tilde{m}-1)-2A_xN_{\tilde{m}-1}\right)^2|\theta_{ji}\right]\right],
\end{aligned}$$

where $N_{\tilde{m}-1}$ denotes the number of occurrences of $-A_x$ in the sequence $x_{ji}^{\tilde{m}-1}$. Given θ_{ji}, the variable $N_{\tilde{m}-1}$ is a binomially distributed random variable with size $(\tilde{m}-1)$ and parameter θ_{ji}. Then we can evaluate $\mathrm{E}[(\sum_{k=1}^{\tilde{m}-1}x_{ji}[i+(k-1)p])^2]$ as

$$\begin{aligned}
&\mathrm{E}\left[\left(\sum_{k=1}^{\tilde{m}-1}x_{ji}[i+(k-1)p]\right)^2\right] \qquad (18)\\
&\qquad = A_x^2\left(2(\tilde{m}-1)\frac{C_j}{2C_j+1}+(\tilde{m}-1)^2\frac{1}{2C_j+1}\right).
\end{aligned}$$

Combining Eqs. (15) to (18) yields for the expected loss of the best sequential regressor in region j

$$\begin{aligned}
&\sum_{m=1}^{n_j}\mathrm{E}\left[(x_j[m]-\hat{x}_{j,q}[m])^2\right] \qquad (19)\\
&= \sum_{i=1}^{p}\sum_{\tilde{m}=1}^{n_{ji}}\left\{A_x^2-2A_x^2\frac{\tilde{m}-1}{(\tilde{m}-1+2C_j)(2C_j+1)}\right.\\
&\qquad + \frac{A_x^2}{(\tilde{m}-1+2C_j)^2}\left(2(\tilde{m}-1)\frac{C_j}{2C_j+1}\right.\\
&\qquad \left.\left. + (\tilde{m}-1)^2\frac{1}{2C_j+1}\right)\right\}.
\end{aligned}$$

3.3. Loss of Batch Regressor

We now proceed and compute the expected loss of the best batch regressor in region j which is given by

$$\underline{w}_j^* = \left(\sum_{k=1}^{n_j}\underline{y}_j[k]\underline{y}_j^{\mathrm{T}}[k]\right)^{-1}\sum_{k=1}^{n_j}\underline{y}_j[k]x_j[k]. \qquad (20)$$

Exploiting the structure of $\underline{y}_j[k]$, the expression $\underline{w}_j^{\mathrm{T}}\underline{y}_j[m]$ can be simplified to

$$\underline{w}_j^{*,\mathrm{T}}\underline{y}_j[m] = \underline{w}_j^{*,\mathrm{T}}\underline{e}_i y_j = \frac{1}{n_{ji}}\sum_{\ell=1}^{n_{ji}}x_j[i+(\ell-1)p], \qquad (21)$$

where again $n_{ji} = \lceil (n_j+1-i)/p \rceil$. Now, the loss of the batch regressor in region j can be written as

$$\begin{aligned}
&\sum_{m=1}^{n_j}\mathrm{E}\left[\left(x_j[m]-\underline{w}_j^{*,\mathrm{T}}\underline{y}_j[m]\right)^2\right] \qquad (22)\\
&= \sum_{i=1}^{p}\sum_{\ell=1}^{n_{ji}}\mathrm{E}\left[\left(x_j[i+(\ell-1)p]-\frac{1}{n_{ji}}\sum_{\ell=1}^{n_{ji}}x_j[i+(\ell-1)p]\right)^2\right].
\end{aligned}$$

As before, $\mathrm{E}[x_{ji}^2[i+(\tilde{m}-1)p]]$ is given by A_x^2. The expectation of the cross term of Eq. (22) can computed as

$$\begin{aligned}
&\mathrm{E}\left[x_{ji}[i+(\tilde{m}-1)p]\sum_{\ell=1}^{n_{ji}}x_{ji}[i+(\ell-1)p]\right] = \qquad (23)\\
&= \mathrm{E}\left[\mathrm{E}\left[x_{ji}^2[i+(\tilde{m}-1)p]|\theta_{ji}\right]+\mathrm{E}\left[x_{ji}[i+(\tilde{m}-1)p]|\theta_{ji}\right]\right.\\
&\qquad \left.\mathrm{E}\left[\sum_{\substack{\ell=1\\ \ell\neq\tilde{m}}}^{n_{ji}}x_{ji}[i+(\ell-1)p]\middle|\theta_{ji}\right]\right]\\
&= \mathrm{E}\left[A_x^2+A_x^2(n_{ji}-1)(1-2\theta_{ji})^2\right] = A_x^2+A_x^2\frac{n_{ji}-1}{2C_j+1}. \qquad (24)
\end{aligned}$$

It remains to compute $\mathrm{E}\left[\left(\sum_{\ell=1}^{n_{ji}}x_{ji}[i+(\ell-1)p]\right)^2\right]$, which can be rewritten using the variable $N_{n_{ji}}$ denoting the number of times that$x_{ji}[i+(\ell-1)p] = -A_x$ in the sequence $x_{ji}^{n_{ji}}$, as

$$\begin{aligned}
&\mathrm{E}\left[\left(\sum_{\ell=1}^{n_{ji}}x_{ji}[i+(\ell-1)p]\right)^2\right] \qquad (25)\\
&\qquad = \mathrm{E}\left[\mathrm{E}\left[\left(n_{ji}A_x-2N_{n_{ji}}A_x\right)^2\middle|\theta_{ji}\right]\right].
\end{aligned}$$

Given θ_{ji}, the distribution of $N_{n_{ji}}$ is binomial with size $n_{ji}$and parameter θ_{ji}, and we get that

$$\begin{aligned}
&\mathrm{E}\left[\left(\sum_{\ell=1}^{n_{ji}}x_{ji}[i+(\ell-1)p]\right)^2\right] \qquad (26)\\
&\qquad = A_x^2\left(2n_{ji}\frac{C_j}{2C_j+1}+n_{ji}^2\frac{1}{2C_j+1}\right).
\end{aligned}$$

Now, the loss of the batch regressor in region j can be written

as

$$\sum_{m=1}^{n_j} \mathrm{E}\left[\left(x_j[m] - \underline{w}_j^{*,\mathrm{T}} \underline{y}_j[m]\right)^2\right]$$
$$= \sum_{i=1}^{p}\sum_{\ell=1}^{n_{ji}} \left\{ A_x^2 - \frac{2A_x^2}{n_{ji}}\left(1 + \frac{n_{ji}-1}{2C_j+1}\right) \right. \quad (27)$$
$$\left. + \frac{A_x^2}{n_{ji}^2}\left(2n_{ji}\frac{C}{2C_j+1} + n_{ji}^2\frac{1}{2C_j+1}\right)\right\}.$$

We can now combine the results obtained in Eqs. (19) and (27) for the expected loss of the best sequential and batch regressor in each region to express $L(n)$, after some simplification, as

$$L(n) = A_x^2 \sum_{j=1}^{J}\sum_{i=1}^{p}\sum_{\tilde{m}=1}^{n_{ji}} \left\{ \frac{2C_j}{(2C_j+1)(\tilde{m}-1+2C_j)} \right. \quad (28)$$
$$\left. + \frac{2C_j}{n_{ji}(2C_j+1)}\right\}.$$

The sum over $\tilde{m}$ is lower bounded by its integral, and yields that

$$L(n)$$
$$\geq A_x^2 \sum_{j=1}^{J}\sum_{i=1}^{p} \frac{2C_j}{(2C_j+1)} \int_{\tilde{m}=0}^{n_{ji}} \frac{1}{\tilde{m}-1+2C_j}\mathrm{d}\tilde{m}$$
$$\geq A_x^2 \sum_{j=1}^{J}\sum_{i=1}^{p} \frac{2C_j}{(2C_j+1)} \left(\ln(n_{ji}-1+2C_j) - \ln(2C_j-1)\right)$$
$$= A_x^2 \sum_{j=1}^{J}\sum_{i=1}^{p} \left\{ \frac{2C_j}{(2C_j+1)} \left(\ln\left(\left\lceil \frac{n_j+1-i}{p}\right\rceil - 1 + 2C_j\right)\right.\right.$$
$$\left.\left. - \ln(2C_j-1)\right)\right\}. \quad (29)$$

Further, we can lower bound n_{ji} as

$$n_{ji} = \left\lceil \frac{n_j+1-i}{p}\right\rceil \geq \frac{n_j+1-i}{p} \geq \frac{n_j+1-p}{p}, \quad (30)$$

since $i \in \{1,2,\ldots,p\}$, and obtain

$$L(n) \geq A_x^2 p \frac{2C}{2C+1}\sum_{j=1}^{J}(n_j) - G, \quad (31)$$

by choosing $C_j = C \geq (2p-1)/(2p)$ for all j and a constant $G = JA_x^2 p\frac{2C}{2C+1}\ln((2C-1)p)$. This lower bound is valid for all integer values n_j satisfying $\sum_{j=1}^{J} n_j = n$. We now let $n_j = \lfloor (n/J)\rfloor$ for $j = 1,2,\ldots,J-1$, and $n_J = n - (J-1)\lfloor (n/J)\rfloor$.

Since $(n/J)-1 \leq \lfloor (n/J)\rfloor \leq (n/J)$, application of Taylor's theorem to $\ln((n/J)-1)$ about (n/J) yields for the lower bound

$$L(n) \geq JA_x^2 p\frac{2C}{2C+1}\left(\ln\left(\frac{n}{J}\right) - \frac{1}{n-J}\right) - G \quad (32)$$
$$\geq JA_x^2 p\frac{2C}{2C+1}\ln\left(\frac{n}{J}\right) - G - \mathrm{O}\left(\frac{1}{n}\right). \quad (33)$$

Hence, for every $0 < \epsilon \leq p/(3p-1)$, we can pick C large enough such that

$$L(n) \geq JA_x^2 p(1-\epsilon)\ln\left(\frac{n}{J}\right) - G - \mathrm{O}\left(\frac{1}{n}\right), \quad (34)$$

which concludes the proof of the theorem.

4. CONCLUSION

Establishing a tight lower bound on the regret of the best sequential regressor with respect to the regressor with access to the entire sequence in advance, we have shown that the piecewise linear regressor presented in [6] is optimal in a min-max sense, in that the regret of any sequential predictor can not be much better.

5. REFERENCES

[1] N. Cesa-Bianchi, P.M. Long, and M.K. Warmuth, "Worst-case quadratic loss bounds for prediction using linear functions and gradient descent," *IEEE Transactions on Neural Networks*, vol. 7, no. 3, pp. 604–619, 1996.

[2] V. Vovk, "Aggregating strategies," in *Proc. 3rd Annual Workshop Comp. Learning Theory*, 1990, pp. 371–383.

[3] V. Vovk, "Competitive online statistics," *International Statistical Review*, vol. 69, pp. 213–248, 2001.

[4] A.C. Singer and M. Feder, "Universal linear prediction by model order weighting," *IEEE Transactions on Information Theory*, vol. 47, no. 10, pp. 2685–2699, October 1999.

[5] A.C. Singer, S.S. Kozat, and M. Feder, "Universal linear least squares prediction: Upper and lower bounds," *IEEE Transactions on Information Theory*, vol. 48, no. 8, pp. 2354–2362, August 2002.

[6] S.S. Kozat, A.C. Singer, and G.C. Zeitler, "Universal piecewise linear prediction via context trees," submitted to IEEE Transactions on Signal Processing.

[7] N. Merhav and M. Feder, "Universal schemes for sequential decision from individual data sequences," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1289–1292, July 1993.