Morvai, G.: Guessing the Output of a Stationary Binary Time Series.

In: Foundations of statistical inference (Shoresh, 2000), 207–215, Contrib. Statist., Physica, Heidelberg, 2003.

Abstract

The forward prediction problem for a binary time series $\{X_n\}_{n=0}^{\infty}$ is to estimate the probability that $X_{n+1} = 1$ based on the observations X_i, $0 \le i \le n$ without prior knowledge of the distribution of the process $\{X_n\}$. It is known that this is not possible if one estimates at all values of n. We present a simple procedure which will attempt to make such a prediction infinitely often at carefully selected stopping times chosen by the algorithm. The growth rate of the stopping times is also exhibited.

1 Introduction

T. Cover in [3] asked two fundamental questions concerning estimation for stationary and ergodic binary processes. Cover's first question was as follows.

Question 1 *Is there an estimation scheme f_{n+1} for the value $P(X_1 = 1|X_0, X_{-1}, \ldots, X_{-n})$ such that f_{n+1} depends solely on the observed data segment $X_0, X_{-1}, \ldots, X_{-n}$ and*

$$\lim_{n\to\infty} f_{n+1}(X_0, X_{-1}, \ldots, X_{-n}) - P(X_1 = 1|X_0, X_{-1}, \ldots, X_{-n}) = 0$$

almost surely for all stationary and ergodic binary time series $\{X_n\}$?.

This question was answered by Ornstein [7] by constructing such a scheme. (See also Bailey [2].) Ornstein's scheme is not a simple one and the proof of consistency is rather sophisticated. A much simpler scheme and proof of consistency were provided by Morvai, Yakowitz, Györfi [6]. (See also Weiss [12].)
Here is Cover's second question.

Question 2 *Is there an estimation scheme f_{n+1} for the value $P(X_{n+1} = 1|X_0, X_1, \ldots, X_n)$ such that f_{n+1} depends solely on the data segment $X_0, X_1, \ldots, X_n$ and*

$$\lim_{n\to\infty} f_{n+1}(X_0, X_1, \ldots, X_n) - P(X_{n+1} = 1|X_0, X_1, \ldots, X_n) = 0$$

almost surely for all stationary and ergodic binary time series $\{X_n\}$*?.*

This question was answered by Bailey [2] in a negative way, that is, he showed that there is no such scheme. (Also see Ryabko [10], Györfi, Morvai, Yakowitz [4] and Weiss [12].) Bailey used the technique of cutting and stacking developed by Ornstein [8] (see also Shields [11]). Ryabko's construction was based on a function of an infinite state Markov-chain. This negative result can be interpreted as follows. Consider a weather forecaster whose task it is to predict the probability of the event 'there will be rain tomorrow' given the observations up to the present day. Bailey's result says that the difference between the estimate and the true conditional probability cannot eventually be small for all stationary weather processes. The difference will be big infinitely often. These results show that there is a great difference between Questions 1 and 2. Question 1 was addressed by Morvai, Yakowitz, Algoet [5] and a very simple estimation scheme was given which satisfies the statement in Question 1 in probability instead of almost surely. Now consider a less ambitious goal than Question 2:

Question 3 *Is there a sequence of stopping times* $\{\lambda_n\}$ *and an estimation scheme* f_n *which depends on the observed data segment* $(X_0, X_1, \ldots, X_{\lambda_n})$ *such that*

$$\lim_{n\to\infty} (f_n(X_0, X_1, \ldots, X_{\lambda_n}) - P(X_{\lambda_n+1} = 1 | X_0, X_1, \ldots, X_{\lambda_n})) = 0$$

almost surely for all stationary binary time series $\{X_n\}$*?*

It turns out that the answer is affirmative and such a scheme will be exhibited below. This result can be interpreted as if the weather forecaster can refrain from predicting, that is, he may say that he does not want to predict today, but will predict at infinitely many time instances, and the difference between the prediction and the true conditional probability will vanish almost surely at the stopping times.

2 Forward Estimation for Stationary Binary Time Series

Let $\{X_n\}_{n=-\infty}^{\infty}$ denote a two-sided stationary binary time series. For $n \geq m$, it will be convenient to use the notation $X_m^n = (X_m, \ldots, X_n)$. For $k = 1, 2, \ldots$, define the sequences $\{\tau_k\}$ and $\{\lambda_k\}$ recursively. Set $\lambda_0 = 0$. Let

$$\tau_k = \min\{t > 0 : X_t^{\lambda_{k-1}+t} = X_0^{\lambda_{k-1}}\}$$

and

$$\lambda_k = \tau_k + \lambda_{k-1}.$$

(By stationarity, the string $X_0^{\lambda_{k-1}}$ must appear in the sequence X_1^{∞} almost surely.) The kth estimate of $P(X_{\lambda_k+1} = 1 | X_0^{\lambda_k})$ is denoted by P_k, and is defined as

$$P_k = \frac{1}{k-1} \sum_{j=1}^{k-1} X_{\lambda_j+1} \qquad (1)$$

For an arbitrary stationary binary time series $\{Y_n\}_{n=-\infty}^0$, for $k=1,2,\ldots$, define the sequence $\hat{\tau}_k$ and $\hat{\lambda}_k$ recursively. Set $\hat{\lambda}_0=0$. Let

$$\hat{\tau}_k=\min\{t>0: Y_{-\hat{\lambda}_{k-1}-t}^{-t}=Y_{-\hat{\lambda}_{k-1}}^0\}$$

and let

$$\hat{\lambda}_k=\hat{\tau}_k+\hat{\lambda}_{k-1}.$$

When there is ambiguity as to which time series $\hat{\tau}_k$ and $\hat{\lambda}_k$ are to be applied, we will use the notation $\hat{\tau}_k(Y_{-\infty}^0)$ and $\hat{\lambda}_k(Y_{-\infty}^0)$.

It will be useful to define another time series $\{\tilde{X}_n\}_{n=-\infty}^0$ as

$$\tilde{X}_{-\lambda_k}^0 := X_0^{\lambda_k} \quad \text{for all } k\geq 1. \tag{2}$$

Since $X_{\lambda_{k+1}-\lambda_k}^{\lambda_{k+1}}=X_0^{\lambda_k}$ the above definition is correct. Notice that it is immediate that $\hat{\tau}_k(\tilde{X}_{-\infty}^0)=\tau_k$ and $\hat{\lambda}_k(\tilde{X}_{-\infty}^0)=\lambda_k$.

Lemma 1 *The two time series $\{\tilde{X}_n\}_{n=-\infty}^0$ and $\{X_n\}_{n=-\infty}^\infty$ have identical distribution, that is, for all $n\geq 0$, and $x_{-n}^0\in\{0,1\}^{n+1}$,*

$$P(\tilde{X}_{-n}^0=x_{-n}^0)=P(X_{-n}^0=x_{-n}^0).$$

Proof First we prove that

$$P(\tilde{X}_{-n}^0=x_{-n}^0,\hat{\lambda}_k(\tilde{X}_{-\infty}^0)=n)=P(X_{-n}^0=x_{-n}^0,\hat{\lambda}_k(X_{-\infty}^0)=n). \tag{3}$$

Indeed, by (2), $\tilde{X}_{-\hat{\lambda}_k(\tilde{X}_{-\infty}^0)}^0=X_0^{\lambda_k}$, and it yields

$$P(\tilde{X}_{-n}^0=x_{-n}^0,\hat{\lambda}_k(\tilde{X}_{-\infty}^0)=n)=P(X_0^n=x_{-n}^0,\lambda_k=n),$$

and by stationarity,

$$P(X_0^n=x_{-n}^0,\lambda_k=n)=P(X_{-n}^0=x_{-n}^0,\hat{\lambda}_k(X_{-\infty}^0)=n)$$

and (3) is proved. Apply (3) in order to get

$$\begin{aligned}
&P(\tilde{X}_{-n}^0=x_{-n}^0)\\
&= \sum_{j=n}^{\infty} P(\tilde{X}_{-n}^0=x_{-n}^0,\hat{\lambda}_n(\tilde{X}_{-\infty}^0)=j)\\
&= \sum_{j=n}^{\infty}\sum_{x_{-j}^{-n-1}\in\{0,1\}^{j-n}} P(\tilde{X}_{-j}^0=x_{-j}^0,\hat{\lambda}_n(\tilde{X}_{-\infty}^0)=j)\\
&= \sum_{j=n}^{\infty}\sum_{x_{-j}^{-n-1}\in\{0,1\}^{j-n}} P(X_{-j}^0=x_{-j}^0,\hat{\lambda}_n(X_{-\infty}^0)=j)\\
&= \sum_{j=n}^{\infty} P(X_{-n}^0=x_{-n}^0,\hat{\lambda}_n(X_{-\infty}^0)=j)\\
&= P(X_{-n}^0=x_{-n}^0)
\end{aligned}$$

and Lemma 1 is proved.

Since $\{X_n\}_{n=-\infty}^{\infty}$ is a stationary time series, by Lemma 1 so is $\{\tilde{X}_n\}_{n=-\infty}^{0}$. Since a stationary time series can always be extended to be a two-sided time series we have also defined $\{\tilde{X}_n\}_{n=-\infty}^{\infty}$. Now we prove the universal consistency of the estimator P_k.

Theorem 1 *For all stationary binary time series $\{X_n\}$ and estimator defined in (1),*

$$\lim_{k\to\infty}\left(P_k - P(X_{\lambda_k+1}=1|X_0^{\lambda_k})\right) = 0 \quad \textit{almost surely.} \tag{4}$$

Moreover,

$$\lim_{k\to\infty} P_k = \lim_{k\to\infty} P(X_{\lambda_k+1}=1|X_0^{\lambda_k}) = P(\tilde{X}_1=1|\tilde{X}_{-\infty}^{0}) \tag{5}$$

almost surely.

Proof

$$\begin{aligned}
& P_k - P(X_{\lambda_k+1}=1|X_0^{\lambda_k}) \\
&= \frac{1}{k-1}\sum_{j=1}^{k-1}\{X_{\lambda_j+1} - P(X_{\lambda_j+1}=1|X_0^{\lambda_j})]\} \\
&+ \frac{1}{k-1}\sum_{j=1}^{k-1}\{P(X_{\lambda_j+1}=1|X_0^{\lambda_j}) - P(X_{\lambda_k+1}=1|X_0^{\lambda_k})\} \\
&= \frac{1}{k-1}\sum_{j=1}^{k-1}\Gamma_j + \frac{1}{k-1}\sum_{j=1}^{k-1}(\Delta_j-\Delta_k).
\end{aligned}$$

Observe that $\{\Gamma_j, \sigma(X_0^{\lambda_j+1})\}$ is a bounded martingale difference sequence for $1\le j<\infty$. To see this note that $\sigma(X_0^{\lambda_j+1})$ is monotone increasing, and Γ_j is measurable with respect to $\sigma(X_0^{\lambda_j+1})$, and $E(\Gamma_j|X_0^{\lambda_{j-1}+1})=0$ for $1\le j<\infty$. Now apply Azuma's exponential bound for bounded martingale differences in Azuma [1] to get that for any $\epsilon>0$,

$$P\left(\left|\frac{1}{(k-1)}\sum_{j=1}^{k-1}\Gamma_j\right| > \epsilon\right) \le 2\exp(-\epsilon^2(k-1)/2).$$

After summing the right side over k, and appealing to the Borel-Cantelli lemma for a sequence of ϵ's tending to zero we get

$$\frac{1}{(k-1)}\sum_{j=1}^{k-1}\Gamma_j \to 0 \quad \text{almost surely.}$$

It remains to show

$$\frac{1}{k-1}\sum_{j=1}^{k-1}\Delta_j - \Delta_k \to 0 \quad \text{almost surely.}$$

Define

$$p_{k,n}(x_{-n}^0) = P(X_{\lambda_k+1} = 1 | X_0^{\lambda_k} = x_{-n}^0, \lambda_k = n)$$

and (applying $\hat{\lambda}_k$ to the time series $\{\tilde{X}_n\}_{n=-\infty}^0$)

$$\tilde{p}_{k,n}(x_{-n}^0) = P(\tilde{X}_1 = 1 | \tilde{X}_{-\hat{\lambda}_k}^0 = x_{-n}^0, \hat{\lambda}_k = n).$$

Now the fact that $\lambda_k = \hat{\lambda}_k$ and Lemma 1 together imply

$$p_{k,n}(x_{-n}^0) = \tilde{p}_{k,n}(x_{-n}^0). \tag{6}$$

By (2) and (6),

$$\tilde{p}_{k,\lambda_k}(X_{\lambda_k}^0) = \tilde{p}_{k,\hat{\lambda}_k}(\tilde{X}_{-\hat{\lambda}_k}^0). \tag{7}$$

Combine (6) and (7) in order to get

$$P(X_{\lambda_k+1} = 1 | X_0^{\lambda_k}) = P(\tilde{X}_1 = 1 | \tilde{X}_{-\hat{\lambda}_k}^0).$$

Notice that $\{P(\tilde{X}_1 = 1 | \tilde{X}_{-\hat{\lambda}_k}^0), \sigma(\tilde{X}_{-\hat{\lambda}_k}^0)\}$ is a bounded martingale and so it converges almost surely to $P(\tilde{X}_1 = 1 | \tilde{X}_{-\infty}^0)$, and so does $P(X_{\lambda_k+1} = 1 | X_0^{\lambda_k})$. We have proved that Δ_j converges almost surely. Now Toeplitz lemma yields that $\frac{1}{k-1}\sum_{j=1}^{k-1}(\Delta_j - \Delta_k) \to 0$ almost surely. The proof of Theorem 1 is complete.

3 The Growth Rate of the Stopping Times

The next result shows that the growth of the stopping times $\{\lambda_k\}$ is rather rapid. Let $p(x_{-n}^0) = P(X_{-n}^0 = x_{-n}^0)$.

Theorem 2 *Let $\{X_n\}$ be a stationary and ergodic binary time series. Suppose that $H > 0$ where*

$$H = \lim_{n\to\infty} -\frac{1}{n+1} E \log p(X_0, \ldots, X_n)$$

is the process entropy. Let $0 < \epsilon < H$ be arbitrary. Then for k large enough,

$$\lambda_k(\omega) \ge c^{c^{\cdot^{\cdot^{c}}}} \text{ almost surely,} \tag{8}$$

where the height of the tower is $k - K$, $K(\omega)$ is a finite number which depends on ω, and $c = 2^{H-\epsilon}$.

PROOF Since by (2), $\lambda_k = \hat{\lambda}_k(\tilde{X}_{-\infty}^0)$, and by Lemma 1 the time series $\{X_n\}_{-\infty}^{\infty}$ and $\{\tilde{X}_n\}_{-\infty}^{\infty}$ have identical distributions, and hence the same entropy, it is enough to prove the result for $\hat{\lambda}_k(\tilde{X}_{-\infty}^0)$. Now $\hat{\tau}_k$ and $\hat{\lambda}_k$ are evaluated on the process $\{\tilde{X}_n\}_{n=-\infty}^0$. For $0 < l < \infty$ define

$$R(l) = \min\{j \ge l+1 : \tilde{X}_{-l-j}^{-j} = \tilde{X}_{-l}^0\}.$$

By Ornstein and Weiss [9],

$$\frac{1}{l+1}\log R(l) \to H \text{ almost surely.} \tag{9}$$

First we show that if $H > 0$ then for k large enough $\hat{\tau}_{k+1} > \hat{\lambda}_k$ almost surely. We argue by contradiction. Suppose that $\hat{\tau}_{k+1} \to \infty$ and $\hat{\tau}_{k+1} \le \hat{\lambda}_k$ infinitely often. Then

$$\tilde{X}^0_{-\hat{\lambda}_k} = \tilde{X}^{-\hat{\tau}_{k+1}}_{-\hat{\lambda}_k - \hat{\tau}_{k+1}}$$

and $\hat{\tau}_{k+1} \le \hat{\lambda}_k$ infinitely often. Hence

$$\tilde{X}^0_{-\hat{\tau}_{k+1}+1} = \tilde{X}^{-\hat{\tau}_{k+1}}_{-\hat{\tau}_{k+1}-\hat{\tau}_{k+1}+1}$$

infinitely often and $R(\hat{\tau}_{k+1} - 1) \le \hat{\tau}_{k+1}$ infinitely often. Then by (9),

$$\begin{aligned} H &= \lim_{k\to\infty} \frac{1}{\hat{\tau}_{k+1}} \log R(\hat{\tau}_{k+1} - 1) \\ &\le \lim_{k\to\infty} \frac{1}{\hat{\tau}_{k+1}} \log \hat{\tau}_{k+1} \\ &= 0 \end{aligned}$$

provided that $\hat{\tau}_k \to \infty$. Now assume that $\eta = \sup_{0<k<\infty} \hat{\tau}_k$ is finite. Then $R(n\eta - 1) = n\eta$. Now by (9),

$$\begin{aligned} H &= \lim_{n\to\infty} \frac{1}{n\eta} \log R(n\eta - 1) \\ &\le \lim_{n\to\infty} \frac{1}{n\eta} \log(n\eta) \\ &= 0. \end{aligned}$$

We have shown that $H > 0$ implies that for k large enough $\hat{\tau}_{k+1} > \hat{\lambda}_k$ almost surely and hence for k large enough $R(\hat{\lambda}_k) = \hat{\tau}_{k+1}$ almost surely. Hence by (9),

$$\frac{1}{\hat{\lambda}_k + 1} \log \hat{\tau}_{k+1} \to H \text{ almost surely.}$$

Thus for almost every $\omega \in \Omega$ there exists a positive finite integer $K(\omega)$ such that for $k \ge K(\omega)$, $\frac{1}{\hat{\lambda}_k+1} \log \hat{\tau}_{k+1} > H - \epsilon$ and

$$\hat{\lambda}_{k+1} > \hat{\tau}_{k+1} > c^{\hat{\lambda}_k} \text{ for } k \ge K(\omega)$$

and the proof of Theorem 2 is complete.

4 Guessing the Output at Stopping Time Instances

If the weather forecaster is pressed to say simply will it rain or not tomorrow then we need a guessing scheme, rather than a predictor. Define the guessing scheme $\{\bar{X}_{\lambda_k}\}$ for the values $\{X_{\lambda_k+1}\}$ as

$$\bar{X}_{\lambda_k} = 1_{\{P_k \geq 0.5\}}.$$

Let $X^*_{\lambda_k}$ denote the Bayes rule, that is,

$$X^*_{\lambda_k} = 1_{\{P(X_{\lambda_k+1}=1|X_0^{\lambda_k}) \geq 0.5\}}.$$

Theorem 3 *Let $\{X_n\}_{n=-\infty}^{\infty}$ be a stationary binary time series. The proposed guessing scheme $\bar{X}_{\lambda_k}$ works in the average at stopping times λ_k just as well as the Bayes rule, that is,*

$$\lim_{n\to\infty} \left(\frac{1}{n} \sum_{k=1}^{n} 1_{\{\bar{X}_{\lambda_k}=X_{\lambda_k+1}\}} - \frac{1}{n} \sum_{k=1}^{n} 1_{\{X^*_{\lambda_k}=X_{\lambda_k+1}\}} \right) = 0 \tag{10}$$

almost surely. Moreover,

$$\lim_{k\to\infty} \left(P(\bar{X}_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) - P(X^*_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) \right) = 0 \tag{11}$$

almost surely.

Proof

$$\begin{aligned}
\sum_{k=1}^{n} 1_{\{\bar{X}_{\lambda_k}=X_{\lambda_k+1}\}} - \frac{1}{n} \sum_{k=1}^{n} 1_{\{X^*_{\lambda_k}=X_{\lambda_k+1}\}} = \\
& \frac{1}{n} \sum_{k=1}^{n} \left[1_{\{\bar{X}_{\lambda_k}=X_{\lambda_k+1}\}} - P(\bar{X}_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) \right] \\
- & \frac{1}{n} \sum_{k=1}^{n} \left[1_{\{X^*_{\lambda_k}=X_{\lambda_k+1}\}} - P(X^*_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) \right] \\
+ & \frac{1}{n} \sum_{k=1}^{n} \left[P(\bar{X}_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) - P(X^*_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) \right] \\
= & \Gamma_n + \Theta_n + \Psi_n.
\end{aligned}$$

Now Γ_n and Θ_n tend to zero since they are averages of bounded martingale differences (cf. Azuma [1]). Concerning the third term Ψ_n, it is enough to prove that

$$\lim_{k\to\infty} \left(P(\bar{X}_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) - P(X^*_{\lambda_k} = X_{\lambda_k+1}|X_0^{\lambda_k}) \right) = 0$$

almost surely. To see this recall the result in Theorem 1,

$$\lim_{k\to\infty} P_k = \lim_{k\to\infty} P(X_{\lambda_k+1} = 1|X_0^{\lambda_k}) = P(\tilde{X}_1 = 1|\tilde{X}^0_{-\infty})$$

almost surely, and apply this in order to get

$$
\begin{aligned}
\lim_{k\to\infty} [P(\bar{X}_{\lambda_k} = X_{\lambda_k+1} | X_0^{\lambda_k}) - P(X^*_{\lambda_k} = X_{\lambda_k+1} | X_0^{\lambda_k})] = \\
\lim_{k\to\infty} \quad & \{[P(P(\tilde{X}_1 = 1 | \tilde{X}^0_{-\infty}) \neq 0.5, \bar{X}_{\lambda_k} = X_{\lambda_k+1} | X_0^{\lambda_k}) \\
- \quad & P(P(\tilde{X}_1 = 1 | \tilde{X}^0_{-\infty}) \neq 0.5, X^*_{\lambda_k} = X_{\lambda_k+1} | X_0^{\lambda_k})] \\
+ \quad & [P(P(\tilde{X}_1 = 1 | \tilde{X}^0_{-\infty}) = 0.5, \bar{X}_{\lambda_k} = X_{\lambda_k+1} | X_0^{\lambda_k}) \\
- \quad & P(P(\tilde{X}_1 = 1 | \tilde{X}^0_{-\infty}) = 0.5, X^*_{\lambda_k} = X_{\lambda_k+1} | X_0^{\lambda_k})]\} \\
= \quad & 0.
\end{aligned}
$$

The proof of Theorem 3 is now complete.

Acknowledgments. The author wishes to thank Benjamin Weiss for helpful discussions and suggestions. This paper has been written by the auspices of the Hungarian National Eötvös Fund. (Ez a cikk a Magyar Állami Eötvös Ösztöndíj támogatásával készült.)

References

[1] K. Azuma, "Weighted sums of certain dependent random variables," in *Tohoku Mathematical Journal*, vol. 37, pp. 357–367, 1967.

[2] D. H. Bailey, *Sequential Schemes for Classifying and Predicting Ergodic Processes.* Ph. D. thesis, Stanford University, 1976.

[3] T. M. Cover, "Open problems in information theory," in *1975 IEEE Joint Workshop on Information Theory*, pp. 35–36. New York: IEEE Press, 1975.

[4] L. Györfi, G. Morvai, and S. Yakowitz, "Limits to consistent on-line forecasting for ergodic time series," *IEEE Transactions on Information Theory*, vol. 44, pp. 886–892, 1998.

[5] G. Morvai, S. Yakowitz, and P. Algoet, "Weakly convergent nonparametric forecasting of stationary time series," *IEEE Transactions on Information Theory*, vol. 43, pp. 483-498, 1997.

[6] G. Morvai, S. Yakowitz, and L. Györfi, "Nonparametric inferences for ergodic, stationary time series," *Annals of Statistics.*, vol. 24, pp. 370–379, 1996.

[7] D. S. Ornstein, "Guessing the next output of a stationary process," *Israel J. Math.*, vol. 30, pp. 292–296, 1978.

[8] D. S. Ornstein, *Ergodic Theory, Randomness, and Dynamical Systems.* Yale University Press, 1974.

[9] D. S. Ornstein and B. Weiss, "Entropy and data compression schemes," *IEEE Transactions on Information Theory*, vol. 39, pp. 78–83, 1993.

[10] B. Ya. Ryabko, "Prediction of random sequences and universal coding," *Problems of Inform. Trans.*, vol. 24, pp. 87-96, Apr.-June 1988.

[11] P.C. Shields, "Cutting and stacking: a method for constructing stationary processes," *IEEE Transactions on Information Theory*, vol. 37, pp. 1605–1614, 1991.

[12] B. Weiss, *Single Orbit Dynamics*, American Mathematical Society, 2000.