Dynamics and time series: theory and applications

Stefano Marmi

Scuola Normale Superiore

Lecture 3, Jan 27, 2009

- Lecture 1: An introduction to dynamical systems and to time series. Periodic and quasiperiodic motions. (Tue Jan 13, 2 pm - 4 pm Aula Bianchi)
- Lecture 2: Ergodicity. Uniform distribution of orbits. Return times. Kac inequality Mixing (Thu Jan 15, 2 pm - 4 pm Aula Dini)
- Lecture 3: Kolmogorov-Sinai entropy. Randomness and deterministic chaos. (Tue Jan 27, 2 pm - 4 pm Aula Bianchi)
- Lecture 4: Time series analysis and embedology. (Thu Jan 29, 2 pm - 4 pm Dini)
- Lecture 5: Fractals and multifractals. (Thu Feb 12, 2 pm - 4 pm Dini)
- Lecture 6: The rhythms of life. (Tue Feb 17, 2 pm - 4 pm Bianchi)
- Lecture 7: Financial time series. (Thu Feb 19, 2 pm - 4 pm Dini)
- Lecture 8: The efficient markets hypothesis. (Tue Mar 3, 2 pm - 4 pm Bianchi)
- Lecture 9: A random walk down Wall Street. (Thu Mar 19, 2 pm - 4 pm Dini)
- Lecture 10: A non-random walk down Wall Street. (Tue Mar 24, 2 pm – 4 pm Bianchi)

- Seminar I: Waiting times, recurrence times ergodicity and quasiperiodic dynamics (D.H. Kim, Suwon, Korea; Thu Jan 22, 2 pm - 4 pm Aula Dini)

- Seminar II: Symbolization of dynamics. Recurrence rates and entropy (S. Galatolo, Università di Pisa; Tue Feb 10, 2 pm - 4 pm Aula Bianchi)

- Seminar III: Heart Rate Variability: a statistical physics point of view (A. Facchini, Università di Siena; Tue Feb 24, 2 pm - 4 pm Aula Bianchi)

- Seminar IV: Study of a population model: the Yoccoz-Birkeland model (D. Papini, Università di Siena; Thu Feb 26, 2 pm - 4 pm Aula Dini)

- Seminar V: Scaling laws in economics (G. Bottazzi, Scuola Superiore Sant'Anna Pisa; Tue Mar 17, 2 pm - 4 pm Aula Bianchi)

- Seminar VI: Complexity, sequence distance and heart rate variability (M. Degli Esposti, Università di Bologna; Thu Mar 26, 2 pm - 4 pm Aula Dini)

- Seminar VII: Forecasting (TBA)

Measure-preserving transformations

X phase space, μ probability measure

Φ:X → **R** observable (a measurable function, say L^2). Let A be subset of X (event).

$\mu(\Phi) = \int_X \Phi \, d\mu$ is the expectation of Φ

T:X→X induces a time evolution

on observables $\Phi \to \Phi \circ T$

on events $A \to T^{-1}(A)$

T is measure preserving if $\mu(\Phi) = \mu(\Phi \circ T)$ i.e. $\mu(A) = \mu(T^{-1}(A))$

Birkhoff theorem and ergodicity

Birkhoff theorem: if T preserves the measure μ then with probability one the time averages of the observables exist (statistical expectations). The system is ergodic if these time averages do not depend on the orbit (statistics and a-priori probability agree)

$$\frac{1}{N}\sum_{0}^{N-1}\varphi\circ T^{i}(x) := \frac{1}{N}S_{N}\varphi(x) \longrightarrow \int_{X}\varphi(t)d\mu(t)$$

$$\frac{1}{N}\#\left\{i\in[0,N), T^{i}(x)\in A\right\} \longrightarrow \mu(A)$$

Law of large numbers: Statistics of orbits = a-priori probability

Recurrence times

- A point is recurrent when it is a point of accumulation of its future (and past) orbit
- Poincarè recurrence: given a dynamical system T which preserves a probability measure μ and a set of positive measure E a point x of E is almost surely recurrent
- First return time of x in E:

  $$R(x,E)=\min\{n>0,\ T^n x \in E\}$$

- E could be an element of a partition of the phase space (symbolic dynamics): this point of view is very important in applications (e.g. the proof of optimality of the Lempel-Ziv data compression algorithm)

Kac's Lemma

- If T is ergodic and E has positive measure then

  $$\int_E R(x,E)d\mu(x)=1 \ ,$$

  i.e. R(x,E) is of the order of 1/μ(E): the average length of time that you need to wait to see a particular symbol is the reciprocal of the probability of a symbol. Thus, we are likely to see the high-probability strings within the window and encode these strings efficiently.

Mixing

Order n correlation coefficient:

$$c_n(\varphi,\psi) := \int \varphi.\psi \circ T^n d\mu - \int \varphi d\mu \int \psi d\mu$$

Ergodicity implies

$$\frac{1}{N}\sum_0^{N-1} c_n(\varphi,\psi) \longrightarrow 0$$

Mixing requires that

$$c_N(\varphi,\psi) \longrightarrow 0$$

namely φ and $\varphi \circ T^n$ become independent of each other as n→∞

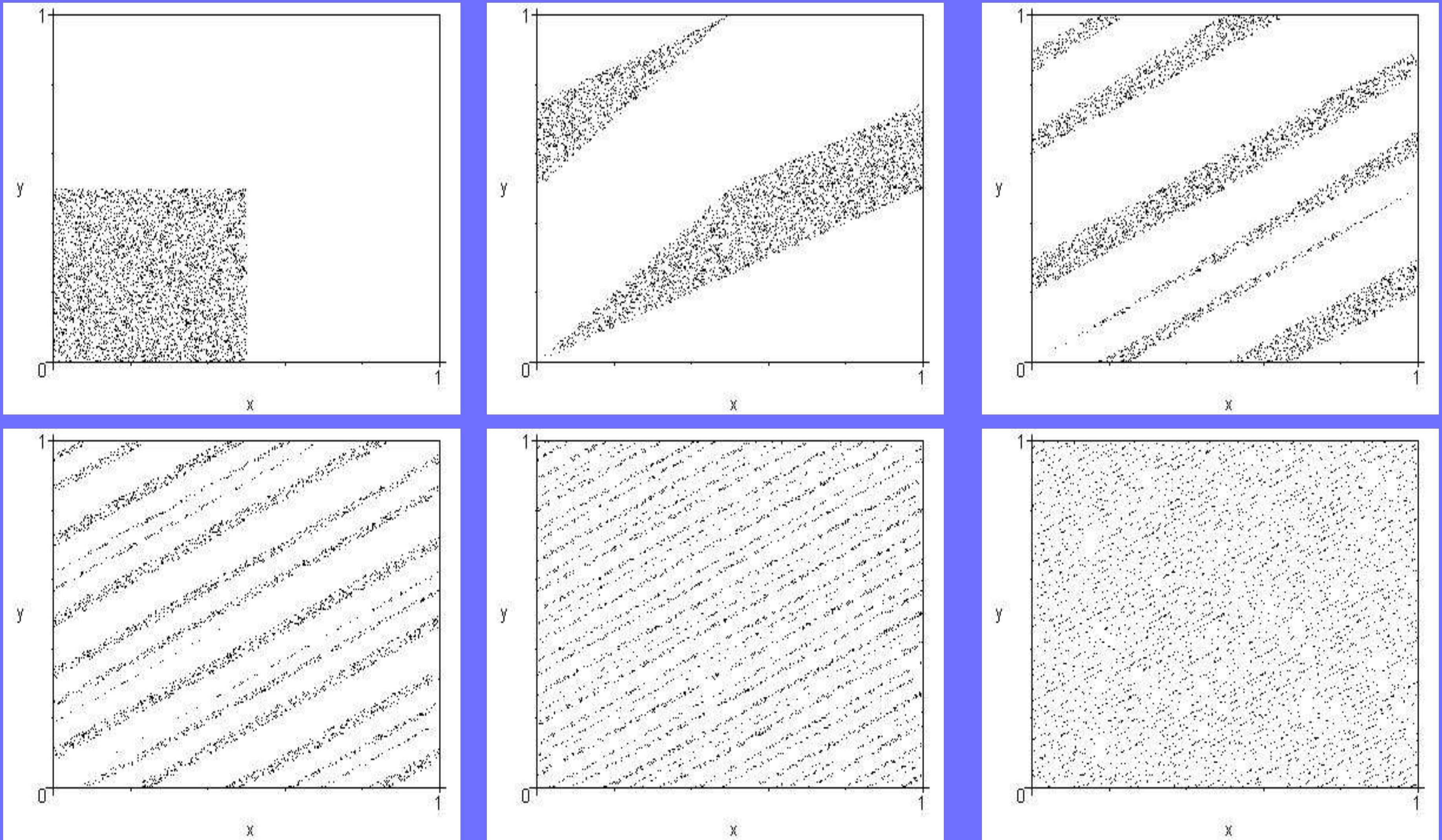
Strong vs. weak mixing

- *Strongly mixing* systems are such that for every E, F, we have $\mu(T^{n}(E)\ \text{п}\ F)\rightarrow \mu\ (E)\ \mu\ (F)$ as n tends to infinity; the Bernoulli shift is a good example. Informally, this is saying that shifted sets become asymptotically independent of unshifted sets.

- *Weakly mixing* systems are such that for every E, F, we have $\mu(T^{n}(E)\ \text{п}\ F)\rightarrow \mu\ (E)\ \mu\ (F)$ as n tends to infinity *after excluding a set of exceptional values of n of asymptotic density zero*.

- Ergodicity does not imply $\mu(T^{n}(E)\ \text{п}\ F)\rightarrow \mu\ (E)\ \mu\ (F)$ but says that this is true for Cesaro averages:

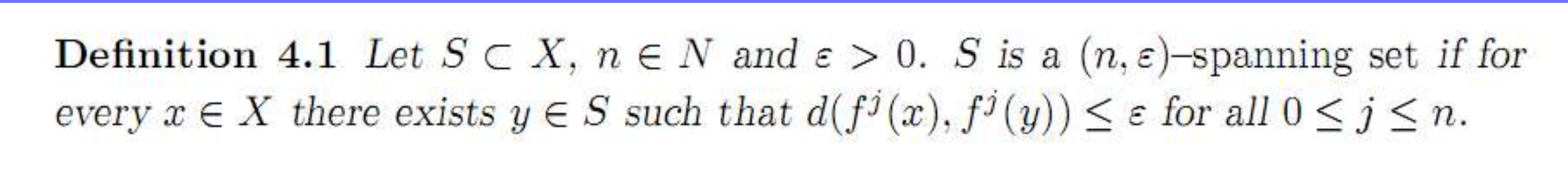
Mixing of hyperbolic automorphisms of the 2-torus (Arnold's cat)



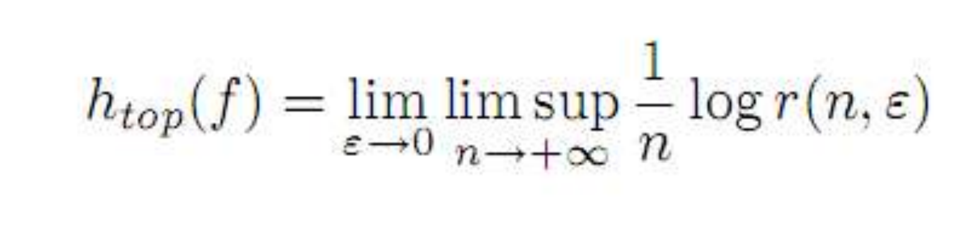
Topological entropy

Topological entropy represents the exponential growth rate of the number of orbit segments which are distinguishable with an arbitrarily high but finite precision. It is invariant under topological conjugacy.

Here the phase space is supposed to be a compact metric space (X.d).



Definition 4.1 *Let $S \subset X$, $n \in N$ and $\varepsilon > 0$. S is a (n,ε)*–spanning set *if for every $x \in X$ there exists $y \in S$ such that $d(f^j(x), f^j(y)) \leq \varepsilon$ for all $0 \leq j \leq n$.*



$$h_{top}(f) = \lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n} \log r(n,\varepsilon)$$

Here r(n,ε) is the least number of points in an (n,ε)-spanning set

Alternative definition

Let α be an open cover of X and let N(α) be the number of sets in a finite subcover of α with smallest cardinality

$$h_{top}(f) = \sup_{\alpha} \lim_{n\to\infty} \frac{1}{n} \log N\left(\bigvee_{i=0}^{n-1} f^{-i}\alpha \right)$$

Here the join αVβ of two covers is obtained considering the sets A∩B where A∊ α, B ∊ β

In information theory, **entropy** is a measure of the uncertainty associated with a random variable.

- Experiment with outcomes $A = \{a_1, \ldots, a_k\}$
- probability of obtaining the result a_i is $p_i, 0 \leq p_i, \leq 1, p_1 + \cdots + p_k = 1$
- If one of the a_i , let us say a_1 occurs with probability that is close to 1, then in most trials the outcome would be a_1 . There is not much information gained after the experiment
- We quantitatively measure the magnitude of 'being surprised' as information = −log (probability)
- (magnitude of our perception is proportional to the logarithm of the magnitude of the stimulus)

Suppose that one performs an experiment which we will denote α which has $m \in \mathbb{N}$ possible mutually esclusive outcomes $A_1, \ldots, A_m$ (e.g. throwing a coin $m = 2$ or a dice $m = 6$). Assume that each possible outcome A_i happens with a probability $p_i \in [0, 1]$, $\sum_{i=1}^m p_i = 1$ (in an experimental situation the probability will be defined statistically). In a probability space $(X, \mathcal{A}, \mu)$ this corresponds to the following setting : α is a finite *partition* $X = A_1 \cup \ldots \cup A_m \bmod(0)$, $A_i \in \mathcal{A}$, $\mu(A_i \cap A_j) = 0$, $\mu(A_i) = p_i$.

Returning to our "experiment", we define on X a function $I(\alpha)$ called *information relative to the partition* α which, evaluated at the point x, expresses the amount of information we get from the knowledge of the element A_i of α to which x belongs. It is natural to ask that I depends only on the probability of A_i so that $I(\alpha) = \sum_{k=1}^m \phi(p_i)\chi_{A_i}$ for some function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$; it is natural to require that ϕ is *decreasing* since the information is bigger if we can locate x in a smaller set. Finally we assume that, if α and β are *independent*, then the information gained from the knowledge of the position of x with respect to both partitions is obtained summing the information relative to each partition : $I(\alpha \vee \beta) = I(\alpha) + I(\beta)$. To fulfill this last requirement on ϕ we must impose that $\phi(ab) = \phi(a) + \phi(b)$ $\forall a, b \in (0, 1)$. It is then clear that $\phi(t)$ must be a constant multiple of $-\log t$.

Entropy and partitions

- Thus the entropy associated to the experiment is

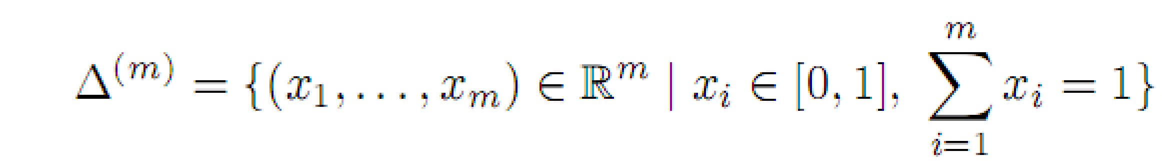
$$H = -\sum_{i=1}^{k} p_i log\, p_i$$

$$\mathcal{P} = \{E_1, \ldots, E_k\}$$

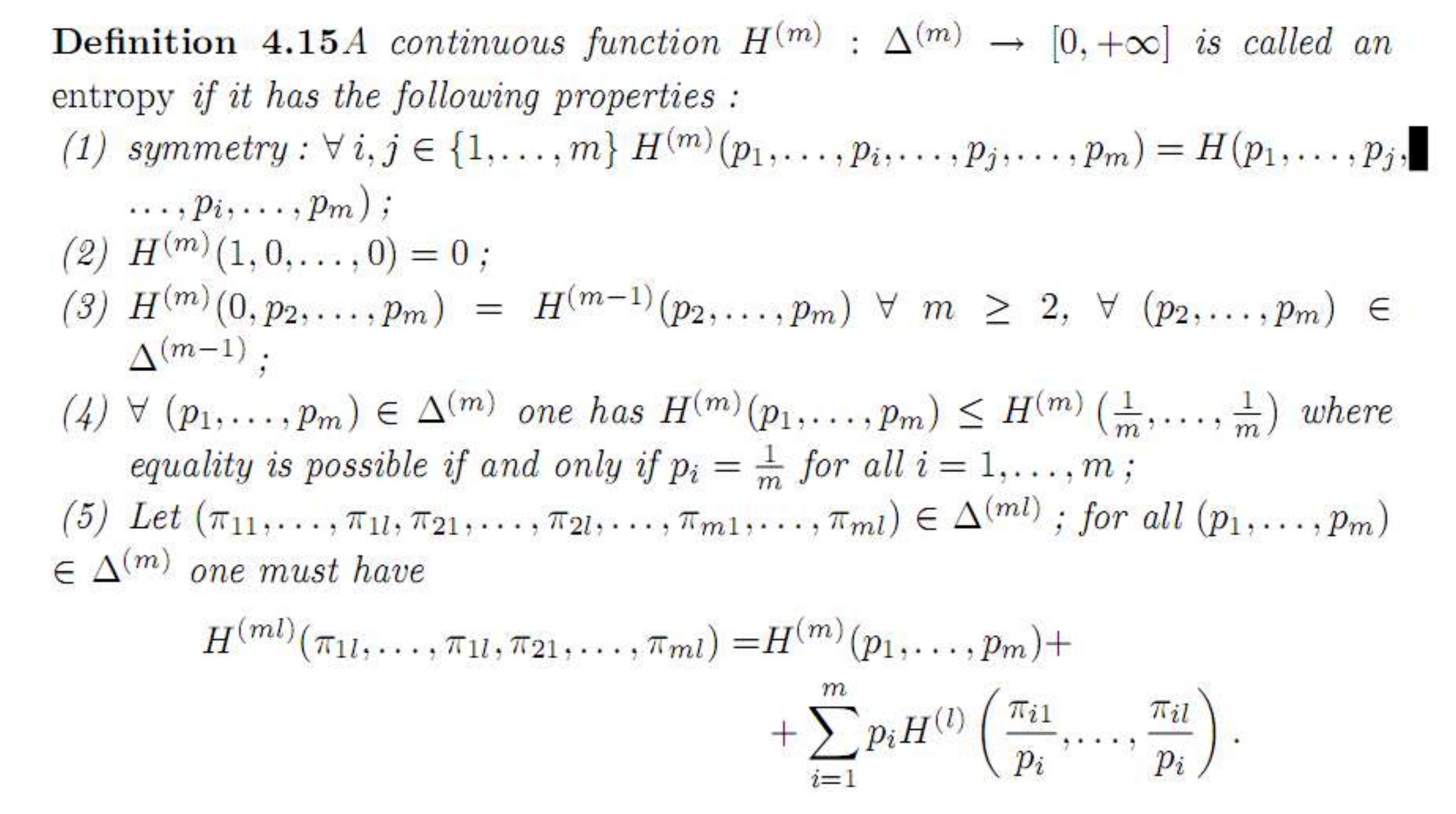
$$p_i = \mu(E_i)$$

In view of the definition of information = - Log (probability), entropy is simply the expectation of information

Uniqueness of entropy



$$\Delta^{(m)} = \{(x_1,\ldots,x_m) \in \mathbb{R}^m \mid x_i \in [0,1],\ \sum_{i=1}^{m} x_i = 1\}$$



Definition 4.15 *A continuous function* $H^{(m)} : \Delta^{(m)} \to [0,+\infty]$ *is called an* entropy *if it has the following properties :*

(1) symmetry : $\forall\, i,j \in \{1,\ldots,m\}$ $H^{(m)}(p_1,\ldots,p_i,\ldots,p_j,\ldots,p_m) = H(p_1,\ldots,p_j,$ $\ldots,p_i,\ldots,p_m)$ *;*

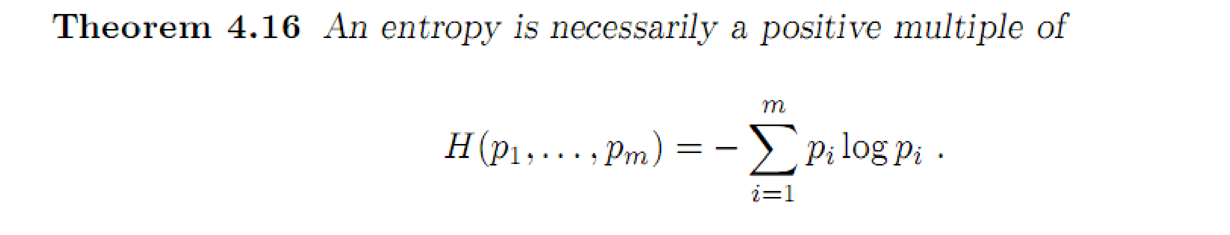
(2) $H^{(m)}(1,0,\ldots,0) = 0$ *;*

(3) $H^{(m)}(0,p_2,\ldots,p_m) = H^{(m-1)}(p_2,\ldots,p_m)$ $\forall\ m \geq 2,\ \forall\ (p_2,\ldots,p_m) \in \Delta^{(m-1)}$ *;*

(4) $\forall\ (p_1,\ldots,p_m) \in \Delta^{(m)}$ *one has* $H^{(m)}(p_1,\ldots,p_m) \leq H^{(m)}\left(\frac{1}{m},\ldots,\frac{1}{m}\right)$ *where equality is possible if and only if* $p_i = \frac{1}{m}$ *for all* $i = 1,\ldots,m$ *;*

(5) Let $(\pi_{11},\ldots,\pi_{1l},\pi_{21},\ldots,\pi_{2l},\ldots,\pi_{m1},\ldots,\pi_{ml}) \in \Delta^{(ml)}$ *; for all* $(p_1,\ldots,p_m)$ $\in \Delta^{(m)}$ *one must have*

$$H^{(ml)}(\pi_{11},\ldots,\pi_{1l},\pi_{21},\ldots,\pi_{ml}) = H^{(m)}(p_1,\ldots,p_m) + \\ + \sum_{i=1}^{m} p_i H^{(l)}\left(\frac{\pi_{i1}}{p_i},\ldots,\frac{\pi_{il}}{p_i}\right).$$



Theorem 4.16 *An entropy is necessarily a positive multiple of*

$$H(p_1,\ldots,p_m) = -\sum_{i=1}^{m} p_i \log p_i\ .$$

Entropy of a dynamical system (Kolmogorov-Sinai entropy)

Given two partitions $\mathcal{P}$ and $\mathcal{Q}$

$\mathcal{P} \vee \mathcal{Q}$ the **join** of $\mathcal{P}$ and $\mathcal{Q}$

$B \cap C$ where $B \in \mathcal{Q}$ and $C \in \mathcal{Q}$

$T : X \to X$ measure preserving

$$\mathcal{P}_n = \mathcal{P} \vee T^{-1}\mathcal{P} \vee \cdots \vee T^{-(n-1)}\mathcal{P}$$

$$h(T, \mathcal{P}) = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}_n)$$

$$h(T) = \sup_{\mathcal{P}} h(T, \mathcal{P})$$

Properties of the entropy

Let $T:X\to X$, $S:Y\to Y$ be measure preserving (T preserves μ, S preserves ν)

$$\textit{If } n \geq 1, \ \textit{then } h(T^n) = n\,h(T)$$

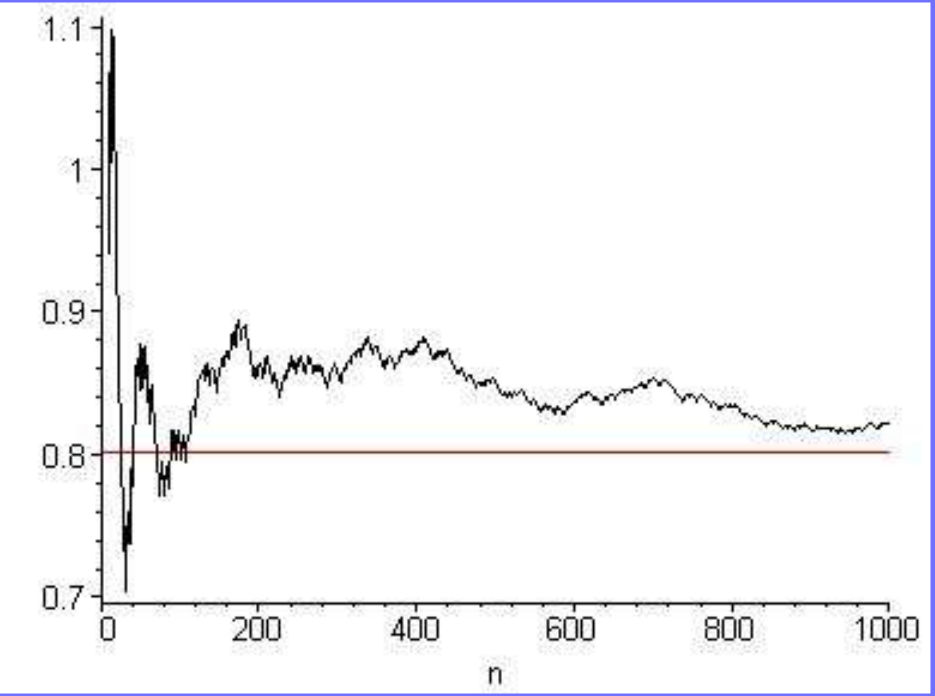
$$\textit{If } T \ \textit{is invertible, then } h(T^{-1}) = h(T)$$

If S is a factor of T then $h(S,\nu)\leq h(T,\mu)$

If S and T are isomorphic then $h(S,\nu)=h(T,\mu)$

On $X\text{x}Y$ one has $h(T\text{x}S,\mu\text{x}\nu)= h(T,\mu)\text{x}h(S,\nu)$

Shannon-Breiman-McMillan



Let $\mathcal{P}$ be a generating partition

Let P(n,x) be the element of

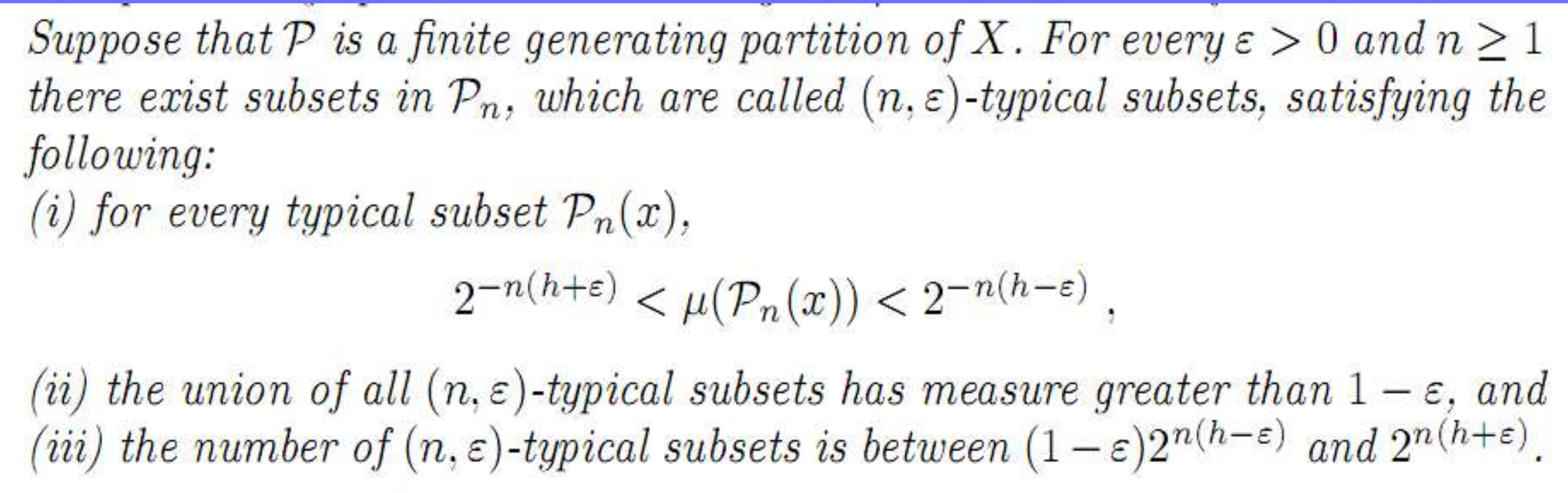
$$\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$$

which contains x

The SHANNON-BREIMAN-MCMILLAN theorem says that for a.e. x one has

$$h(T,\mu) = -\lim_{n\to\infty} \frac{\mathrm{Log}\ \mu(P(n,x))}{n}$$

Asymptotic equipartition property



Suppose that $\mathcal{P}$ is a finite generating partition of X. For every $\varepsilon > 0$ and $n \geq 1$ there exist subsets in $\mathcal{P}_n$, which are called (n, ε)-typical subsets, satisfying the following:

(i) for every typical subset $\mathcal{P}_n(x)$,

$$2^{-n(h+\varepsilon)} < \mu(\mathcal{P}_n(x)) < 2^{-n(h-\varepsilon)} ,$$

(ii) the union of all (n, ε)-typical subsets has measure greater than $1 - \varepsilon$, and

(iii) the number of (n, ε)-typical subsets is between $(1-\varepsilon)2^{n(h-\varepsilon)}$ and $2^{n(h+\varepsilon)}$.

These formulas assume that the entropy is measured in bits, i.e. using the base 2 logarithm

Entropy of Bernoulli schemes

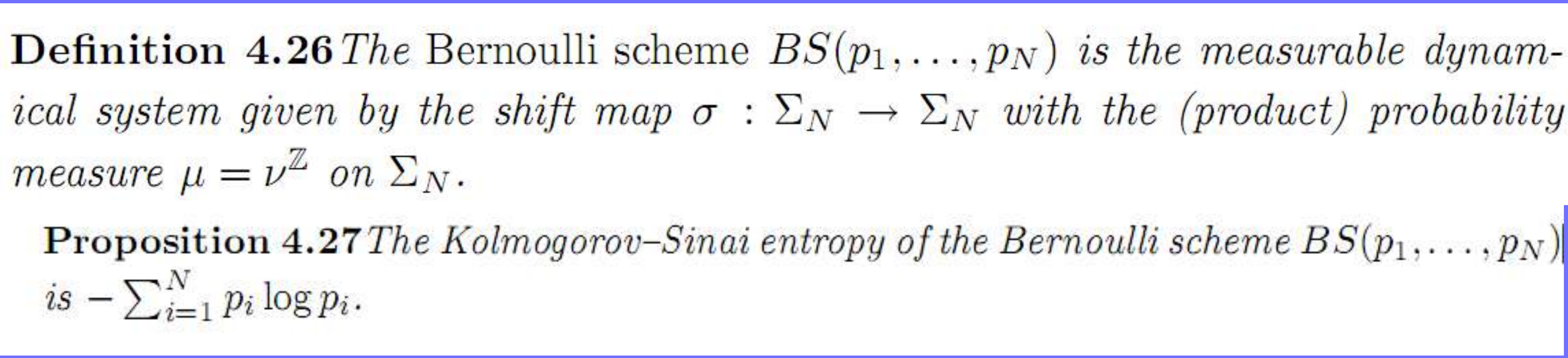
Let $N \geq 2$, $\Sigma_N = \{1, \dots N\}^{\mathbb{Z}}$.

$$d(x,y) = 2^{-a(x,y)} \quad \text{where } a(x,y) = \inf\{|n|\,,\, n \in \mathbb{Z}\,,\, x_n \neq y_n\}$$

shift $\sigma \ : \Sigma_N \to \Sigma_N$ $\qquad \sigma((x_i)_{i\in\mathbb{Z}}) = (x_{i+1})_{i\in\mathbb{Z}}$

The topological entropy of (Σ_N, σ) is $\log N$

$(p_1, \dots, p_N) \in \Delta^{(N)}$ $\qquad \nu(\{i\}) = p_i$



Definition 4.26 *The* Bernoulli scheme $BS(p_1, \dots, p_N)$ *is the measurable dynamical system given by the shift map* $\sigma \ : \Sigma_N \to \Sigma_N$ *with the (product) probability measure* $\mu = \nu^{\mathbb{Z}}$ *on* Σ_N.

Proposition 4.27 *The Kolmogorov–Sinai entropy of the Bernoulli scheme* $BS(p_1, \dots, p_N)$ *is* $-\sum_{i=1}^{N} p_i \log p_i$.

Lyapunov exponent for a map of an interval

- Assume that T is a piecewise smooth map of I=[0,1]
- By the chain rule we have

$$\frac{1}{n}\log|T^n(x)-T^n(y)| \approx \frac{1}{n}\sum_{i=0}^{n-1}\log|T'(T^ix)| \ .$$

- If μ is an ergodic invariant measure for a.e. x the limit exists and it is given by $\int_0^1 \log|T'|\,\mathrm{d}\mu$ it is also called the Lyapunov exponent of T

Expanding maps and Rokhlin formula

If T is expanding then it has a unique a.c.i.p.m. μ and the entropy h of T w.r.t. μ is equal to the Lyapunov exponent

$$h = \int_0^1 \log |T'(x)| \, \mathrm{d}\mu$$

Topological Markov chains or subshifts of finite type

$\Sigma_A = \{x \in \Sigma_N \,, (x_i, x_{i+1}) \in \Gamma \, \forall i \in \mathbb{Z}\}$

$\Gamma \subset \{1, \ldots N\}^2$

Σ_A is a compact shift invariant subset of Σ_N

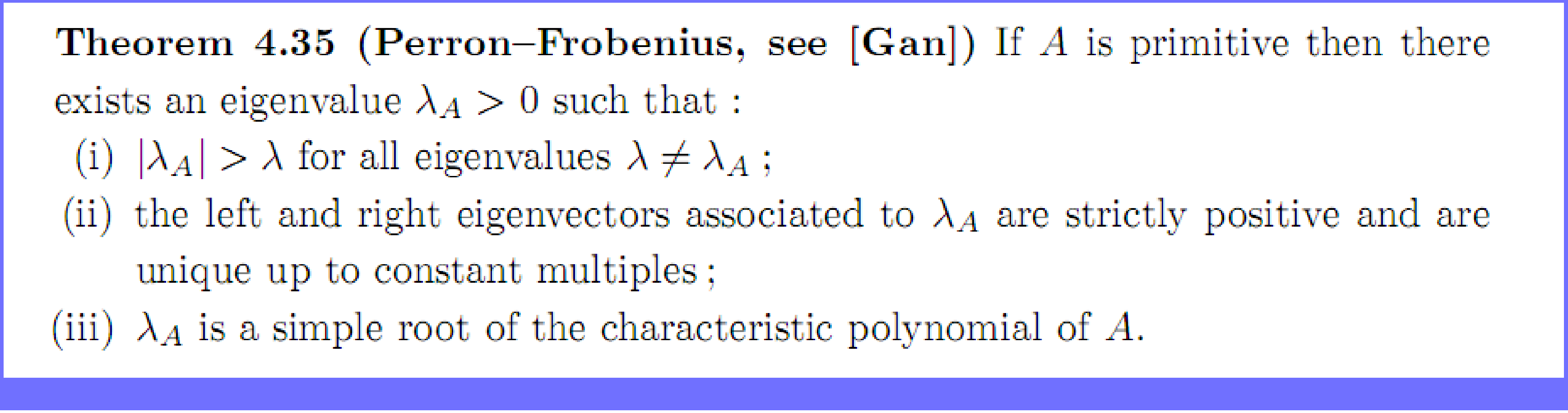
$A = A_\Gamma$ the $N \times N$ matrix with entries $a_{ij} \in \{0, 1\}$

$$a_{ij} = \begin{cases} 1 & \iff (i, j) \in \Gamma \\ 0 & \text{otherwise} \end{cases}$$

The restriction of the shift σ to Σ_A is denoted σ_A

$A^m = (a^m_{ij})$ and $a^m_{ij} > 0$ for all i, j

Entropy of Markov chains

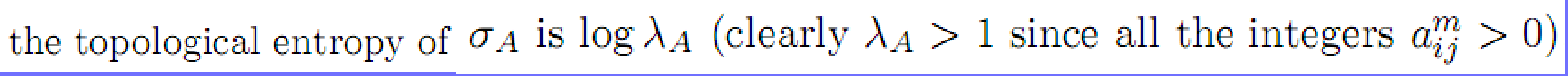


Theorem 4.35 (Perron–Frobenius, see [Gan]) If A is primitive then there exists an eigenvalue $\lambda_A > 0$ such that :

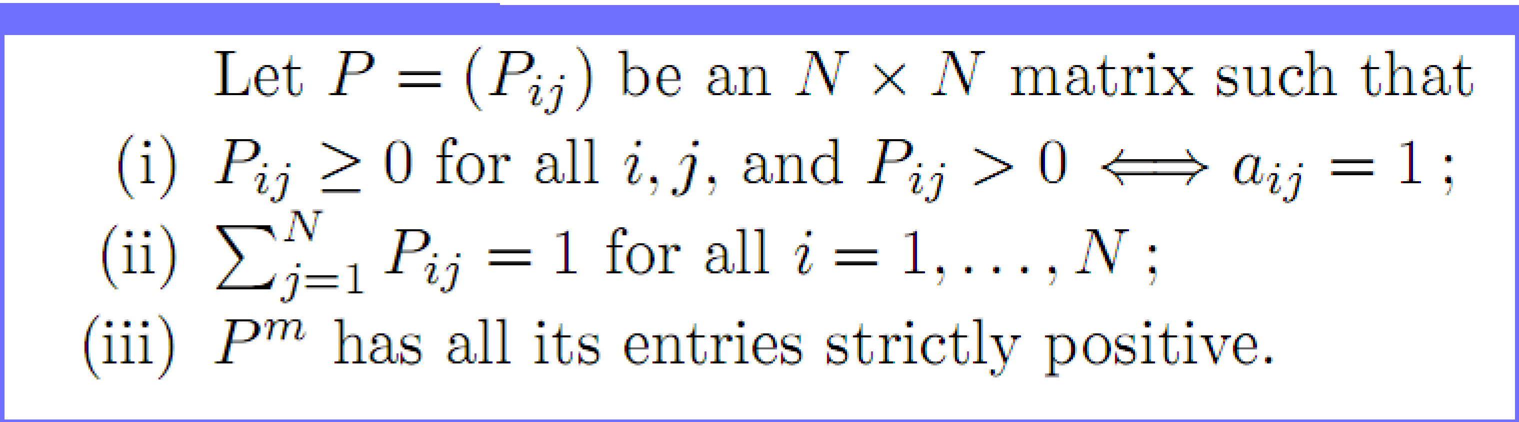
(i) $|\lambda_A| > \lambda$ for all eigenvalues $\lambda \neq \lambda_A$;

(ii) the left and right eigenvectors associated to λ_A are strictly positive and are unique up to constant multiples ;

(iii) λ_A is a simple root of the characteristic polynomial of A.



the topological entropy of σ_A is $\log \lambda_A$ (clearly $\lambda_A > 1$ since all the integers $a_{ij}^m > 0$)



Let $P = (P_{ij})$ be an $N \times N$ matrix such that

(i) $P_{ij} \geq 0$ for all i, j, and $P_{ij} > 0 \iff a_{ij} = 1$;

(ii) $\sum_{j=1}^{N} P_{ij} = 1$ for all $i = 1, \ldots, N$;

(iii) P^m has all its entries strictly positive.

Such a matrix is called a *stochastic matrix*. Applying Perron–Frobenius theorem to P we see that 1 is a simple eigenvalue of P and there exists a normalized eigenvector $p = (p_1, \ldots, p_N) \in \Delta^{(N)}$ such that $p_i > 0$ for all i and

$$\sum_{i=1}^{N} p_i P_{ij} = p_j \ , \quad \forall\, 1 \leq i \leq N \ .$$

We define a probability measure μ on Σ_A corresponding to P prescribing its value on the cylinders :

$$\mu\left(C\begin{pmatrix} j_0, \ldots, j_k \\ i, \ldots, i+k \end{pmatrix}\right) = p_{j_0} P_{j_0 j_1} \cdots P_{j_{k-1} j_k} \ ,$$

for all $i \in \mathbb{Z}$, $k \geq 0$ and $j_0, \ldots, j_k \in \{1, \ldots, N\}$. It is called the *Markov measure* associated to the stochastic matrix P.

the subshift σ_A preserves the Markov measure μ.

$$h_\mu(\sigma_A) = -\sum_{i,j=1}^{N} p_i P_{ij} \log P_{ij}$$

$$h_\mu(\sigma_A) \leq h_{top}(\sigma_A)$$

Entropy, coding and data compression

- Computer file= infinitely long binary sequence
- Entropy = best possible compression ratio
- Lempel-Ziv (Compression of individual sequences via variable rate coding, IEEE Trans. Inf. Th. 24 (1978) 530-536): does not assume knowledge of probability distribution of the source and achieves asymptotic compression ratio=entropy of source

Let $X = \{0,1\}^{\mathbb{N}}$ and σ be a left-shift map.
Define R_n to be the first return time of the initial n-block, i.e.,

$$R_n(x) = \min\{j \geq 1 : x_1 \ldots x_n = x_{j+1} \ldots x_{j+n}\}.$$

$$x = \overbrace{\boxed{1\,0\,1\,0}\,0\,1\,0\,0\,1\,1\,0\,1\,1\,0\,0}^{15}\boxed{1\,0\,1\,0}\cdots \Rightarrow R_4(x) = 15.$$

The convergence of $\dfrac{1}{n}\log R_n(x)$ to the entropy h was studied in a relation with data compression algorithm such as the Lempel-Ziv compression algorithm.

The Lempel-Ziv data compression algorithm provide a universal way to coding a sequence without knowledge of source.
Parse a source sequence into shortest words that has not appeared so far:

$$1011010100010\cdots \Rightarrow 1, 0, 11, 01, 010, 00, 10, \ldots$$

For each new word, find a phrase consisting of all but the last bit, and recode the location of the phrase and the last bit as the compressed data.

$$(000, 1)\,(000, 0)\,(001, 1)\,(010, 1)\,(100, 0)\,(010, 0)\,(001, 0)...$$

Theorem (Wyner-Ziv(1989), Ornstein and Weiss(1993))

For ergodic processes with entropy h,

$$\lim_{n\to\infty} \frac{1}{n} \log R_n(x) = h \quad \textit{almost surely.}$$

The meaning of entropy

- Entropy measures the information content or the amount of randomness.
- Entropy measures the maximum compression rate.
- Totally random binary sequence has entropy $\log 2 = 1$. It cannot be compressed further.

The entropy of English

Is English is a stationary ergodic process? Probably not!

Stochastic approximations to English: as we increase the complexity of the model, we can generate text that looks like English. The stochastic models can be used to compress English text. The better the stochastic approximation, the better the compression.

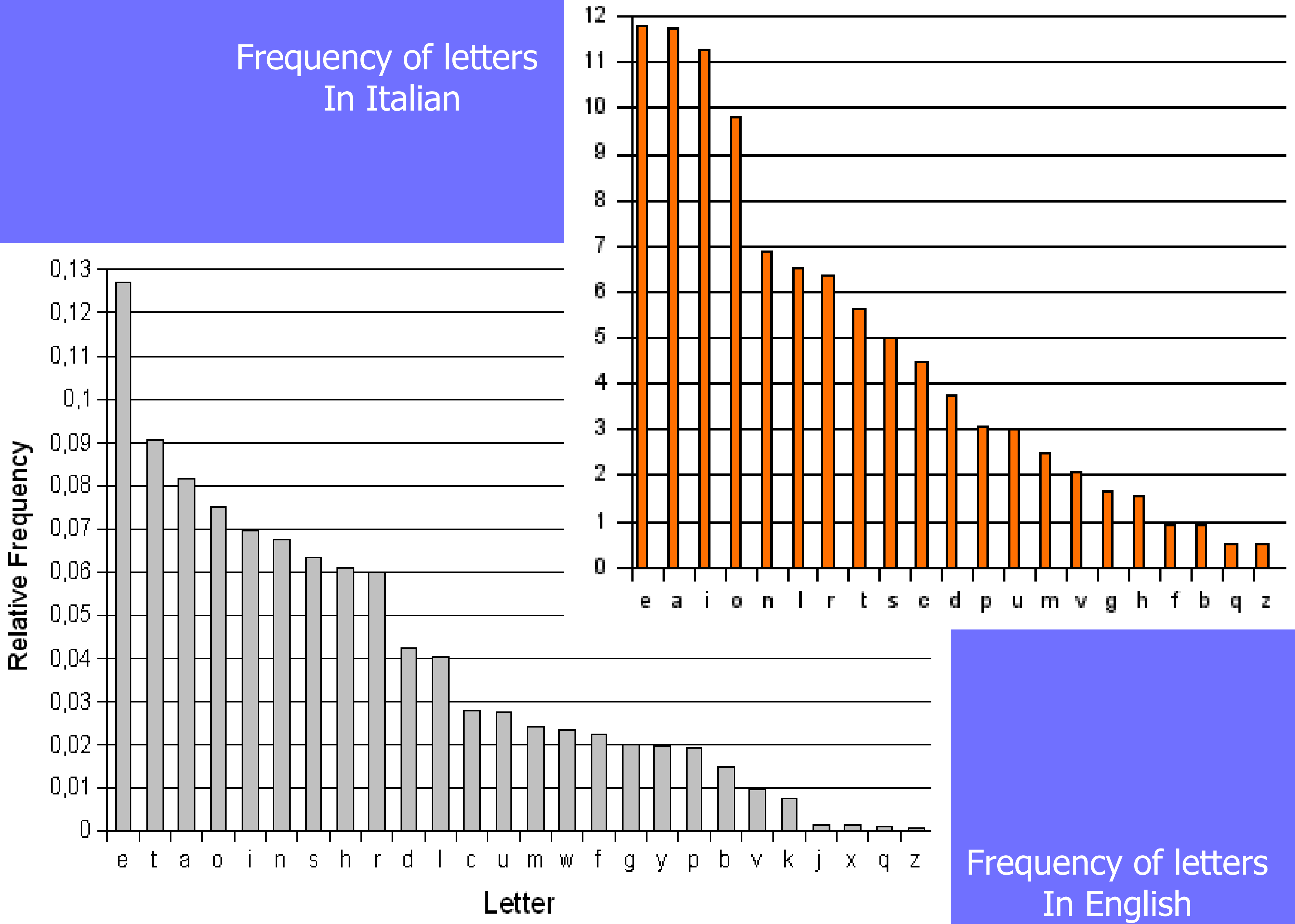
alphabet of English = 26 letters and the space symbol

models for English are constructed using empirical distributions collected from samples of text.

E is most common, with a frequency of about 13%,

least common letters, Q and Z, have a frequency of about 0.1%.

Frequency of letters In Italian



Frequency of letters In English

Source: Wikipedia

Construction of a Markov model for English

The frequency of pairs of letters is also far from uniform:

Q is always followed by a U, the most frequent pair is TH,

(frequency of about 3.7%), etc.

Proceeding this way, we can also estimate higher-order conditional probabilities and build more complex models for the language.

However, we soon run out of data. For example, to build

a third-order Markov approximation, we must compute

$p(xi |xi-1,xi-2,xi-3)$ in correspondence of $27x27^3 = 531\ 441$ entries for this table: need to process millions of letters to make accurate estimates of these probabilities.

Examples (Cover and Thomas, Elements of Information Theory, 2nd edition , Wiley 2006)

- Zero order approximation (equiprobable h=4.76 bits):

  XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD

- First order approximation (frequencies match):

  OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI

  ALHENHTTPA OOBTTVA NAH BRL

- Second order (frequencies of pairs match): ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE
- Third order (frequencies of triplets match): IN NO IST LAT WHEY CRATICT FROURE BERS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE

- Fourth order approximation (frequencies of quadruplets match, each letter depends on previous three letters; h=2.8 bits):

  THE GENERATED JOB PROVIDUAL BETTER TRANDTHE DISPLAYED CODE, ABOVERY UPONDULTS WELL THE CODERST IN THESTICAL IT DO HOCK BOTHE MERG. (INSTATES CONS ERATION. NEVER ANY OF PUBLE AND TO THEORY. EVENTIAL CALLEGAND TO ELAST BENERATED IN WITH PIES AS IS WITH THE)
- First order WORD approximation (random words, frequencies match): REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.
- Second order (WORD transition probabilities match): THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED