BERNOULLI SHIFT **Math118, O. Knill**

ABSTRACT. When equiped with an invariant measure, which is the area measure when representing it as the Baker map, the shift is called the Bernoulli shift. It produces independent random variables.

A SHIFT INVARIANT MEASURE. We have defined a map S from the unit square $Y = [0,1) \times [0,1)$ to the sequence space $X = \{0,1\}^Z$ by

$$S(u,v)_n = \begin{cases} 0 & u_n < 1/2 \\ 1 & u_n \geq 1/2 \end{cases}$$

if $T^n(u,v) = (u_n, v_n)$ is the orbit of the Baker map. This was called **symbolic dynamics**. We can use the map S to measure subsets in X by requiring that it preservs the measure: the left half of the square of area $1/2$ is maped into the set of sequences x which satisfy $x_0 = 0$, the right half of the square of area $1/2$ is maped into $\{x \mid x_0 = 1\}$. The set $\{x_0 = 0, x_1 = 1\}$ in X corresponds to the lower left quarter of the square which has area $1/4$.

THE BERNOULLI MEASURE. The space X can be equiped with a shift invariant probability measure P. In that case, we say $P[U]$ is the measure or the probability of U. We can define $P[U]$ as the area of $S^{-1}(U)$ in the square. We know then that

$$P[x_{n+1} = f_1, ..., x_{n+m} = f_m] = 2^{-m} .$$

This measure is called a **Bernoulli measure**. It is **invariant under the shift**. for any subset U of X, then $P[\sigma(U)] = P[U]$.

If U is a subset of the square and S is the map conjugating the Baker map to the shift, then $P[S(U)]$ is the area of U.

RANDOM VARIABLE. A **random variable** is a (continuous) function from X to R. Examples of random variables are $X_k(x) = x_k$. Two random variables are called **independent** if P$[\{Y = a, Z = b\}]$ = P$[\{Y = a\}]$P$[\{Z = b\}]$ for any choice a, b.

The random variables $X_k = x_k$ in the Bernoulli shift are independent.

PROOF. P$[X_k = a, X_l = b] = 1/4$ for any choice of a, b. This is the same as P$[Y = a]$P$[Z = b] = (1/2) \cdot (1/2)$.

In other words, one can use the Bernoulli shift or the Baker map to produce **random numbers**.
This is not a very practical way to produce random numbers: lets look at the first coordinates, when applying the Baker map, we have $T^n(x) = 2^n x$ mod 1. If we start with a a rational number, then $T^n(x)$ will be attracted by a periodic orbit like for example $1/3, 2/3, 1/3, ...$ For a practical generation of random numbers other maps are better suited.

EXPECTATION. The **expectation** of a random variable which takes finitely many values $f_1, ..., f_m$ is

$$E[Y] = P[Y = f_1]f_1 + ... + P[Y = f_n]f_n$$

Two random variables Y, Z are called **uncorrelated** if $E[YZ] = E[Y]E[Z]$. Two independent random variables are automatically uncorrelated.

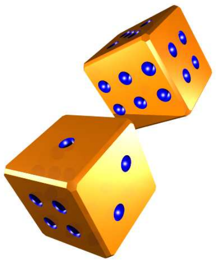
EXAMPLE. $A = \{\text{head}, \text{tail}\}$ models throwing a coin The random variable

$$X(x) = \left[\begin{array}{ll} 3 & x_0 = \text{head} \\ 5 & x_1 = \text{tail} \end{array} \right.$$

has the expectation

$$E[X] = P[X = 3]3 + P[X = 5]5 = 3/2 + 5/2 = 4 .$$

EXAMPLE. Consider the shift over the alphabet $A = \{1, 2, 3, ..., 6\}$. The random variables $X_1, X_2, ...$ simulate the outcomes of a dice event. If $X_3 = 5$, then the third dice rolling produced a 5. These random variables are uncorrelated and independent.



THE LAW OF LARGE NUMBERS. The law of large numbers tells that if X_k are independent random variables with the same distribution, then

$$\frac{1}{n} \sum_{k=1}^{n} X_k$$

converges to the common expectation $E[X_k]$ for almost all experiments.

EXAMPLES. In the dice case, we have for almost all sequences x, that $\frac{1}{n} \sum_{k=1}^{n} X_k \to 7/2$.

OTHER MEASURES. The set X can be equiped with other measures. Assume the letter $x_k = 1$ should have probability p and $x_k = 0$ should have probability $1 - p$. In that case, the probability P$[x_1 = a_1, ..., x_n = a_n]$ is $\binom{n}{k} p^k (1-p)^{n-k}$, where k is the number of times, $a_i = 1$. Knowing the probability of all these events defines the invariant measure. All theses measures are called Bernoulli measures.

MARKOV CHAINS. Often, one does not know the invariant measure, but one knows the conditional probabilities: $P[x_{n+1} = a | x_n = b] = M_{ab}$. In words, the probability that $x_{n+1} = a$ under the condition $x_n = b$ is P_{ab}. The matrix M_{ab} is called a **Markov matrix**. It has the property that the sum of coefficients in each column is equal to 1. The matrix M is a $n \times n$ matrix, if the alphabet A has n elements. You have seen examples of the following fact in linear algebra:

The eigenvector $p = (p_1, ..., p_n)$ to the eigenvalue q of the matrix M normalized so that the $p_1 + ... + p_n = 1$ defines a Bernoulli probability measure on X.

EXAMPLES.
a) If $M_{ij} = 1/2$ for all i, j, we have the Bernoulli shift.
b) If $M = \begin{bmatrix} 1/2 & 2/3 \\ 1/2 & 1/3 \end{bmatrix}$, we can read off the probabilities p that $x_n = 1$ and $1 - p$ that $x_n = 0$ by computing the eigenvector v of M to the eigenvalue 1 and normalizing it, so that the sum of its entries is 1.

MEASURES ON SUBSHIFTS OF FINITE TYPE. If we use a Markov matrix for which $M_{ab} = 0$ if ab is a forbidden word, then we obtain an invariant measure for the subhift of finite type. Subshifts of finite type have a lot of invariant measures. Markov matrices provide a possibility to define such measures.

MEASURES ON SUBSHIFTS. Every subshift X has an invariant measure. It can be obtained by averaging along an orbit. This averaging does not converge in general, but there is a subsequence, along which the limit sets $P[A] = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} 1_{T^k(x) \in A}$

UNIQUELY ERGODIC SUBSHIFTS. If there is only one shift-invariant measure then the subshift is called **uniquely ergodic**. An example are Sturmean sequences, which are obtained by doing symbolic dynamics on using a half open I and an irrational rotation on the circle. The reason that there is only one invariant measure is because also the irrational rotation on the circle has only one invariant measure.

ERGODIC THEORY. The part of dynamical systems, which deals with invariant measures of a map or dynamical system is called **ergodic theory**. It has close relations to probability theory. The law of large numbers we mentioned here has a generalization which is called **Birkhoffs ergodic theorem**.