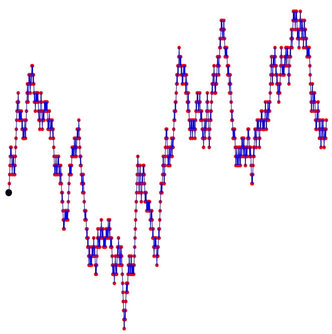
Lecture 31: The law of large numbers

A sequence of random variables is called IID abbreviating **independent, identically distributed** if they all have the same distribution and if they are independent.

We assume that all random variables have a finite variance $\mathrm{Var}[X]$ and expectation $\mathrm{E}[X]$.

A sequence of random variables defines a **random walk** $S_n = \sum_{k=1}^n X_k$. The interpretation is that X_k are the individual steps. If we take n steps, we reach S_n.

Here is a typical trajectory of a random walk. We throw a dice and if the dice shows head we go up, if the dice shows tail, we go down.



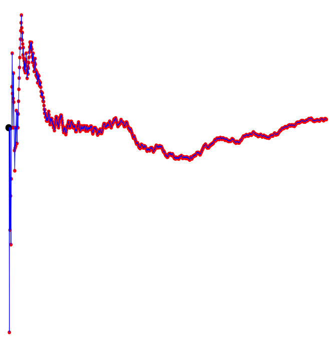
The following result is one of the three most important results in probability theory:

Law of large numbers. For almost all ω, we have $S_n/n \to \mathrm{E}[X]$.

Proof. We only prove the weak law of large numbers which deals with a weaker convergence: We have $\mathrm{Var}[S_n/n] = n\mathrm{Var}[X]/n^2 = \mathrm{Var}[X]/n$ so that by Chebyshev's theorem

$$\mathrm{P}[|S_n/n - \mathrm{E}[X]| < \epsilon] \leq \mathrm{Var}[X]/n\epsilon^2$$

for $n \to \infty$. We see that the probability that S_n/n deviates by a certain amount from the mean goes to zero as $n \to \infty$. The strong law would need a half an hour for a careful proof. [1]



[1] O.Knill, Probability and Stochastic Processes with applications, 2009

1 If X_i are random variables which take the values $0, 1$ and 1 is chosen with probability p, then S_n has the binomial distribution and $\mathrm{E}[S_n] = np$. Since $\mathrm{E}[X] = p$, the law of large numbers is satisfied.

2 If X_i are random variables which take the value $1, -1$ with equal probability $1/2$, then S_n is a symmetric random walk. In this case, $S_n/n \to 0$.

3 Here is a strange paradox called the **Martingale paradox**. We try it out in class. Go into a Casino and play the doubling strategy. Enter 1 dollar, if you lose, double to 2 dollars, if you lose, double to 4 dollars etc. The first time you win, stop and leave the Casino. You won 1 dollar because you lost maybe 4 times and $1+2+4+8 = 15$ dollars but won 16. The paradox is that the expected win is zero and in actual Casinos even negative. The usual solution to the paradox is that as longer you play as more you win but also increase the chance that you lose huge leading to a zero net win. It does not quite solve the paradox because in a Casino where you are allowed to borrow arbitrary amounts and where no bet limit exists, you can not lose.

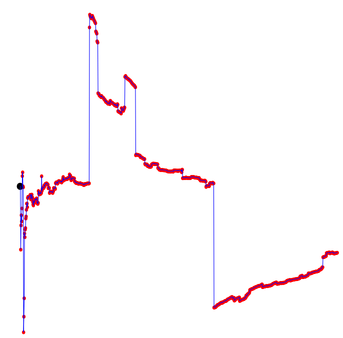
How close is S_n/n to $\mathrm{E}[X]$? Experiment:

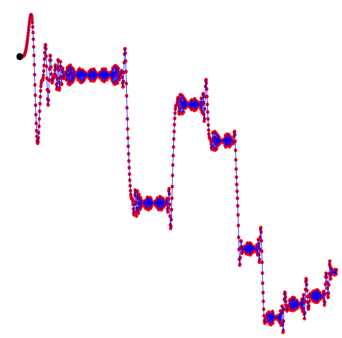
4 Throw a dice n times and add up the total number S_n of eyes. Estimate $S_n/n - \mathrm{E}[X]$ with experiments. Below is example code for Mathematica. How fast does the error decrease?

```
f[n_]:=Sum[Random[Integer,5]+1,{n}]/n-7/2;
data=Table[ {k,f[k]},{k,1000}]; Fit[data,{1,Exp[-x]},x]
```

5 Here is the situation where the random variables are Cauchy distributed. The expectation is not defined. The left picture below shows this situation.

6 What happens if we relax the assumption that the random variables are uncorrelated? The illustration to the right below shows an experiment, where we take a periodic function $f(x)$ and an irrational number α and where $X_k(x) = f(k\alpha)$.





It turns out that no randomness is necessary to establish the strong law of large numbers. It is enough to have "ergodicity"

A probability preserving transformation T on a probability space (Ω, P) is called **ergodic** if every event A which is left invariant has probability 0 or 1.

7 If Ω is the interval $[0,1]$ with measure $\mathrm{P}[[c,d]] = d-c$, then $T(x) = x + \alpha \bmod 1$ is ergodic if α is irrational.

Birkhoff's ergodic theorem. If $X_k = f(T^k x)$ is a sequence of random variables obtained from an ergodic process, then $S_n(\omega)/n \to \mathrm{E}[X]$ for almost all ω.

This theorem is the reason that ideas from probability theory can be applied in much more general contexts, like in **number theory** or in **celestial mechanics**.

Application: normal numbers

A real number is called **normal** to base 10 if in its decimal expansion, every digit appears with the same frequency 1/10.

Almost every real number is normal

The reason is that we can look at the k'th digit of a number as the value of a random variable $X_k(\omega)$ where $\omega \in [0,1]$. These random variables are all independent and have the same distribution. For the digit 7 for example, look at the random variables $Y_k(\omega) = \begin{cases} 1 & \omega_k = 7 \\ 0 & else \end{cases}$ which have expectation 1/10. The average $S_n(\omega)/n =$ "number of digits 7 in the first k digits of the decimal expansion" of ω converges to 1/10 by the law of large numbers. We can do that for any digit and therefore, almost all numbers are normal.

Application: Monte Carlo integration

The limit

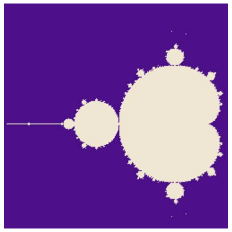
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(x_k)$$

where x_k are IID random variables in $[a,b]$ is called the **Monte-Carlo integral**.

The Monte Carlo integral is the same than the Riemann integral for continuous functions.

We can use this to compute areas of complicated regions:

The following two lines evaluate the **area of the Mandelbrot fractal** using Monte Carlo integration. The function F is equal to 1, if the parameter value c of the quadratic map $z \to z^2 + c$ is in the Mandelbrot set and 0
8 else. It shoots 100'000 random points and counts what fraction of the square of area 9 is covered by the set. Numerical experiments give values close to the actual value around 1.51.... One could use more points to get more accurate estimates.



```
F[c_]:=Block[{z=c,u=1},Do[z=N[z^2+c];If[Abs[z]>3,u=0;z=3],{99}];u];
M=10^5; Sum[F[-2.5+3 Random[]+I(-1.5+3 Random[])],{M}]*(9.0/M)
```

Application: first significant digits

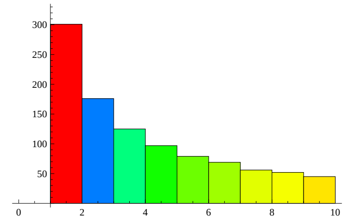
If you look at the distribution of the first digits of the numbers 2^n: $2,4,8,1,3,6,1,2,\ldots$. Lets experiment:

```
data=Table[First[IntegerDigits[2^n]],{n,1,100}];
Histogram[data,10]
```

Interestingly, the digits are not equally distributed. The smaller digits appear with larger probability. This is called **Benford's law** and it is abundant. Lists of numbers of real-life source data are distributed in a nonuniform way. Examples are bills, accounting books, stock prices. Benfords law stats the digit k appears with probability

$$p_k = \log_{10}(k+1) - \log_{10}(k)$$

where $k = 1, \ldots, 9$. It is useful for **forensic accounting** or investigating **election frauds**.



The probability distribution p_k on $\{1, \ldots, 9\}$ is called the **Benford distribution**.

The reason for this distribution is that it is uniform on a logarithmic scale. Since numbers x for which the first digit is 1 satisfy $0 \leq \log(x) \bmod 1 < \log_{10}(2) = 0.301\ldots$, the chance to have a digit 1 is about 30 percent. The numbers x for which the first digit is 6 satisfy $0.778\ldots = \log_{10}(6) \leq \log(x) \bmod < \log_{10}(7) = 0.845\ldots$, the chance to see a 6 is about 6.7 percent.

Homework due April 20, 2011

1 Look the first significant digit X_n of the sequence 2^n. For example $X_5 = 3$ because $2^5 = 32$. To which number does S_n/n converge? We know that X_n has the Benford distribution $\mathrm{P}[X_n = k] = p_k$. [The actual process which generates the random variables is the irrational rotation $x \to x + \log_{10}(2) \bmod 1$ which is ergodic since the $\log_{10}(2)$ is irrational. You can therefore assume that the law of large numbers (Birkhoffs generalization) applies.]

2 By going through the proof of the weak law of large numbers, does the proof also work if X_n are only uncorrelated?

3 Assume A_n is a sequence of $n \times n$ upper-triangular random matrices for which each entry is either 1 or 2 and 2 is chosen with probability $p = 1/4$.
a) What can you say about $\mathrm{tr}(A_n)/n$ in the limit $n \to \infty$?
b) What can you say about $\log(\det(A_n))/n$ in the limit $n \to \infty$?