PROBABILITY

In recent years there has been a marked increase in the number of graduate students specializing in probability as well as an increase in the number of professors who work in this area. This recent boom in the interest in the subject of probability makes it essential that first year graduate students become aware of the basic concepts and interesting problems surrounding this area. On March 30^{th}, April 6^{th}, and April 13^{th} lectures were given by professors Mohammud Foondun, Firas Rassoul-Agha, Davar Khoshnevisan, and graduate student Karim Khader in order to impart the importance and power of probability for the Early Reasearch Directions class. Transcribing these lectures were Matthew Housley, William Malone, and Joshua Shrader.

Mohammud Foondun

Stochastic Processes and Brownian Motion: The first real work in this area was done by Louis Bachelier around 1900.

Definition 1. *A real valued stochastic process* $\{B(t) : t > 0\}$, $x \in \mathbb{R}$ *is a Brownian motion if it satisfies*

(1) $B(0) = x$ *(x is the starting point.)*
(2) *The process has independent increments*

$$0 < t_0 < x_1 < \cdots < t_n$$

such that

$$B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}), \ldots, B(t_1) - B(t_0)$$

are independent.
(3) *For each $h > 0$, the function*

$$B(t+h) - B(t)$$

is a normal distribution with mean 0 and variance h.
(4) *The map $t \mapsto B(t)$ is continuous a.s. (almost surely).*

Almost surely is equivalent to measure theoretic definition of almost everywhere in an event space.

Properties of Brownian motion:

(1) Nowhere differentiable (almost surely).
(2) Total variation is infinite.

Markov Property: The function $B(t+s) - B(t)$ does not depend on what happens before time t.

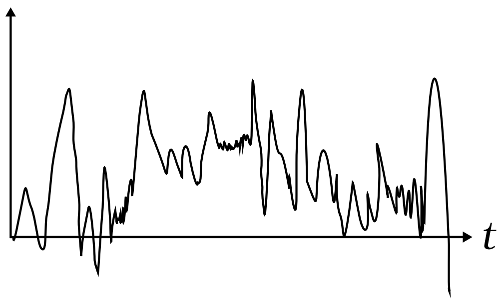

FIGURE 1. A graph of typical one-dimensional Brownian motion.

Technical Definition:
If

$$[f(B(t+s)) \mid \mathcal{F}_s^0] = \mathbb{E}^B(s) f(B_L),$$

then

$$\mathcal{F}_s^0 = \vartheta(B_L, t \leq s)$$

does not depend on what happened before time t. ($\mathcal{F}_s^0$ is filtration).

Levy Process: (generalization of above)
X_t is a Levy process if for all $s, t \geq 0$ the increment $X_{t+s} - X_t$ is independent of $(X_L : 0 \leq L \leq t)$ and it has the same law as X_s.

Levy-Khintchine:

$$\mathbb{E}e^{iz\chi_t} = e^{t\Psi(z)},$$

where $\mathbb{E}$ is the expectation and $\Psi(z)$ is the characteristic exponent, which is given by

$$\Psi(z) = \underbrace{-\frac{1}{2} z \cdot Az}_{\text{continuous}} + i\eta \cdot z + \underbrace{\int_{\mathbb{R}^d} (e^{iz \cdot x} - 1 - uz \cdot x) \sim (dx)}_{\text{jump part}}.$$

Drift is given by η and A is a matrix.

$$\int_{\mathbb{R}^d} |x|^2 \wedge \mid \nu \; d(x) < \infty.$$

The symbol $\wedge$ indicates the minimum of two quantities and (A, r, χ) is a Levy Triplet.
Generation of a Levy Process

$$P_t F(x) = \mathbb{E}[F(X + tX)].$$

The generator L of the process is defined as

$$LF(x) = \lim_{t \to 0} \frac{P_t F - F}{t}, \; F \in C_0^2,$$

i.e.,

$$LF(x) = \frac{1}{2}\sum A_{i,j}\partial_{i,j}F(x) + \sum \chi_i \partial_i F(x)$$
$$+ \int_{\mathbb{R}^d} [F(x+h) - F(x) - \sum h_i \frac{\partial F}{\partial x_i}\mathcal{I}_{|n|\leq 1}]\tau(dh).$$

(We need Fourier transforms and their inverses to derive this equation.) One notes that

$$LF(x) = \frac{1}{2}\partial_{i,j}F(x)$$

if A is the identity matrix and the jump part is zero.

Martingale Problem of Strouck and Varndham:
Martingales: $(r, \mathcal{F}_t, \mathbb{P})$

(1) X_t is adapted to $\mathcal{F}_t$.
(2) $\mathbb{E}|X_t| < \infty$
(3) $\mathbb{E}[X_s/\mathcal{F}_t] = X_t$.

This means that the expectation at any future time is given by your current position. A probability measure $\mathbb{P}$ is a solution to the Martingale problem of L started at x if $\mathbb{P}(x_0 = x) = 1$ and $F(x_t) - F(x_0) - \int_0^t LF(x_s)ds$ is a $\mathbb{P}^x$ (local Martingale). Therefore,

$$LF(x) = a_{i,j}(x)\partial_{i,j}F(x) + b_i(x)\partial_i F(x)$$
$$LF(x) = \int [F(x+h) - F(x)]n(x,h)dh.$$

L is the generator of this process (via handwaviness). If L is a third derivative, then there is no process.

Ito's Lemma:
Let $F \in C_0^2(\mathbb{R}^d)$, and B_τ a Brownian motion.

$$F(B_\tau) - F(B_0) = \int_0^t F(B_s)dB_s + \frac{1}{2}\int_0^t F''(B_s)ds\langle B\rangle_s.$$

$\langle B\rangle_t$ is the quadratic variation of B, i.e. $B_t^2 - \langle B\rangle_t$ whould be a Martingale.

Dirichlet Problem:
Let D be a ball, $Lu = 0$ in D, $u = F$ on ∂D, $u \in C^2$. Then,

$$u(x) = \mathbb{E}F(X_{\tau_D}),$$
$$L = \frac{1}{2}\Delta u,$$
$$I_D = \inf\{t > 0, X_t \notin D\}.$$

Proof. Let $S_n = \inf\{t > 0, \operatorname{dist}(x_t, \partial D) < \frac{1}{n}\}$.

$$u(x_t \wedge S_n) = u(x_0) + M_t + \int_0^{t\wedge S_n} Lu(x_s)ds.$$

Take the expectation

$$\mathbb{E}^x u(x_{t\wedge S^n}) = u(x).$$

Letting t and n go to infinity we get

$$u(x) = \mathbb{E}(x_{\tau_0}).$$

The solution is the average along the boundary. □

Firas Rassoul-Agha

A random walk may be thought of as a sequence of independent identically distributed (iid) random variables (X_i) defined on the d-dimensional lattice of integers $\mathbb{Z}^d$. The probability of moving from one point to any adjacent point is $\frac{1}{2d}$, since there are $2d$ adjacent points. The probability of moving from a point to any non-adjacent point is 0. Define a random walk, $S_n = X_1 + X_2 + \cdots + X_n$ for $n \geq 1$, and $S_0 = 0$. For any arbitrary function $f : \mathbb{Z}^d \to \mathbb{R}$, we define

$$u(n,x) = \mathbb{E}\left[f(x+S_n)\right]$$

as a function in time and space which returns the expected value of f. S_n is a random walk which starts at the origin, and so $x + S_n$ is the walk S_n starting at x. We may re-write the above equation as

$$u(n,x) = \sum_{|e|=1} \mathbb{E}\left[f(x+S_{n-1} + eI\left\{x_n = e\right\}\right],$$

where I is the indicator function, or characteristic function, and is defined to be 1 if $x_n = e$ and 0 otherwise. Since the probability that $X_n = e$ is $\frac{1}{2d}$, we have

$$u(n,x) = \frac{1}{2d}\sum_{|e|=1} \mathbb{E}\left[f(x+e+S_{n-1})\right] = \frac{1}{2d}\sum_{|e|=1} u(n-1, x+e).$$

From this, we can compute the time differential, $\partial_t u$.

$$\partial_t u = u(n,x) - u(n-1,x) = \frac{1}{2d}\sum_{|e|=1}\left(u(n-1,x+e) - u(n-1,x)\right) = \frac{1}{2d}\Delta u.$$

This is the discrete version of the derivative.

If we restrict our attention to the 1 dimensional case and resale time and space, Brownian motion will result. Scaling time by $1/n$ and space by $1/\sqrt{n}$, we have

$$u(t,x) = \mathbb{E}\left[f\left(x + \frac{S_{[nt]}}{\sqrt{n}}\right)\right]$$

which follows from the Central Limit Theorem. This theorem says that $\frac{X_1+X_2+\cdots+X_n}{\sqrt{n}}$ follows the normal distribution $N(0, dI)$. Thus,

$$\lim_{n\to\infty} u(t,x) = \mathbb{E}\left[f(x+\sqrt{t}dZ\right] = \sqrt{d}Z\left(\frac{1}{\sqrt{2\pi t d}}\right)^{d-1}\int f(y)e^{\frac{-|y-1|^2}{2td}}\,\mathrm{d}y$$

and so $u(t,x)$ exhibits Brownian motion.

Consider the following question concerning random walks. Let D be the annulus given by $D = \{x : r_1 < |x| < r_2\}$. Given a random walk starting at some $x \in D$, what is the probability that the random walk will intersect the outer boundary of the annulus, $B_{r_2}(0)$, before intersecting the inner boundary, $R_{r_1}(0)$?

To answer this question, define σ_{r_2} to be the time at which the random walk reaches the boundary at r_2, and define σ_{r_1} similarly. Let $\Delta u = 0$ and $u = I_{|x|=R}$ if $|x| \in \{r_1, r_2\}$. Define

$$u(x) = \frac{\varphi(|x|) - \varphi(r)}{\varphi(R) - \varphi(r)},$$

where

$$\varphi(s) = \begin{cases} s & d = 1 \\ \log(s) & d = 2 \\ s^{2-d} & d \geq 3. \end{cases}$$

Now, define $\tau x = \inf\{n \geq 0 : S_n + x \in \partial D\}$, and so

$$\mathbb{E}_x\left[I_{|B_\tau|=R}\right] = P_x\left(\sigma_{r_2} < \sigma_{r_1}\right).$$

$P_x\left(\sigma_{r_2} < \sigma_{r_1}\right) = \frac{\varphi(|x|)-\varphi(r)}{\varphi(R)-\varphi(r)}$, and so in the 1-dimensional case, we have $P_x\left(\sigma_{r_2} < \sigma_0\right) = |x|/r_2$. Thus, as $r_2 \to \infty$, we see that $Px(\sigma_0 = \infty) = 0$ and so for all y, $P_0(\sigma_y = \infty) = 0$. In other words, this tells us that 1-dimensional Brownian motion hits all points with probability 1.

If we consider the same question in dimension 2, we see that

$$P_x\left(\sigma_{r_2} < \sigma_{r_1}\right) = \frac{\log|x| - \log r_1}{\log r_2 - \log r_1}.$$

As $r_2 \to \infty$, $P_x\left(\sigma_{r_1=\infty}\right) = 0$, and so 2-dimensional Brownian motion will hit any open region with probability 1, but fixing a single point, with probability 1, the Brownian motion will not hit it.

As the final case, consider when $d = 3$. We then have

$$P_x(\sigma_{r_2} < \sigma_{r_1}) = \frac{|x|^{2-d} - r_1^{2-d}}{r_2^{2-d} - r_1^{2-d}}.$$

As $r_2 \to \infty$, $P_x(\sigma_{r_2} < \sigma_{r_1}) = 1 - |x|^{2-d}/r^{2-d} < 1$, and so we can propagate the estimate $P_0(|B_\tau| \to \infty) = 1$. This says that 3-dimensional Brownian motion always wanders off to infinity. In other words, a drunken sailor will always wander back home; however, with positive probability, the drunken parrot will never make it back.

To add a new level of complexity to our random walks, consider the case where the probability of moving to one point instead of another is based on your current position. In one dimension, this can be thought of as laying down coins at all the integer points. Starting at the origin, pick up the coin and flip it. If it lands with heads facing up, move to the left. If it lands with tails facing up, move to the right. If all of the coins are fair, then you have a 50% chance of moving left, and a 50% chance of moving right. However, what if we have two different coins, one red and one blue, each biased in their own way. These red and blue coins are randomly distributed along all of the integer points. If we start at 0, we may ask what the probability is that we will wander to infinity. This question has a physical analog. If one considers an electron jumping from atom to atom in a pure crystal, this will represent a truly random walk; however, if there are impurities in the crystal, then the decisions that the electron makes will be based on its current position. This is analogous to the situation where we have two biased coins.

To solve this problem, we first fix the set of coins on the integers, (P_x). Let a and b be two integers, and define

$$V(x) = P_x(\sigma_b < \sigma_a).$$

Then

$$(P_x + 1 - P_x)\,V(x) = P_x V(x+1) + (1 - P_x)V(x-1)$$

and so

$$V(x+1) - V(x) = \frac{1 - P_x}{P_x}\left(V(x) - V(x-1)\right) = R_x I,$$

where $R_x = \Pi_{y=a+1}^{x} \frac{1-P_y}{P_y}$, and I is some constant. From the above equation, it becomes apparent why V, R, and I were chosen as they were, as this is simply an expression of Ohm's Law – The voltage drop between two points is equal to the resistance between those points times the current. Since

$$1 = V(b) - V(a) = RI,$$

we have that $I = 1/R$ and

$$V(x+1) - V(x) = \frac{\Pi_{y=a+1}^{x} \frac{1-P_y}{P_y}}{\sum_{z=a}^{b-1} \Pi_{y=1}^{z} \frac{P_y}{1-P_y}}.$$

This gives

$$P_x(\tau_b < \tau_a) = \frac{\sum_{z=a}^{x-1} \Pi_{y=a}^{z} \frac{P_y}{1-P_y}}{\sum_{z=a}^{b-a} \Pi_{y=a}^{z} \frac{P_y}{1-P_y}}.$$

Now, letting $b \to \infty$, we may ask what the probability is that we hit a. We have

$$P_x(\tau_a = \infty) = \frac{\sum_{z=a}^{x-1} \Pi_{y=a}^{z} \frac{P_y}{1-P_y}}{\sum_{z=a}^{\infty} \Pi_{y=a}^{z} \frac{P_y}{1-P_y}},$$

where the denominator is either almost surely infinite for all (P_x) or finite for all (P_x). Taking the logarithm of the denominator, we get

$$\sum_{y=a}^{z} \log \frac{P_y}{1 - P_y} \sim (z - a)\,\mathbb{E}\left[\log \frac{P}{1 - P}\right].$$

Now, the law of large numbers tells us that the expected value of $\log \frac{P}{1-P}$ is less than 0 if and only if $P(S_n \to \infty) = 1$ for almost every (P_x).

As an example of a random walk in a non-random environment, consider the following case. There are two coins, red and blue. The red coin tells you to move left 40% of the time, and tells you to move right 60% of the time. The blue coin tells you to move left 80% of the time, and tells you to move right 20% of the time. You begin your walk at the origin, and each time you reach a new vertex, you flip a decision coin. The decision coin tells you to flip the red coin 90% of the time, and to flip the blue coin 10% of the time. The expected value of $\log \frac{P}{1-P}$ is thus

$$\mathbb{E}\left(\log \frac{P}{1-P}\right) = .9 \log \frac{3}{2} + .1 \log \frac{1}{4} \approx .2263 \ .$$

Since this expected value is positive, we will eventually walk to positive infinity. If the value were negative, we would walk to negative infinity, and if this expected value were zero, we could conclude that the walk would be recurrent – we would come back to the origin infinitely often.

Karim Khader

Let (Ω, A, P) be a measure space satisfying $P(\Omega) = 1$. Let $\Omega = [0, 1)$. The collection A is the Borel σ-algebra on Ω and P is Lebesgue measure. A real valued random variable is a measurable function $F : \Omega \to \mathbb{R}$, $\mathbb{E}[F] = \int_\Omega F(\omega)P(d\omega)$. Let X be a measurable subset of Ω. Then, $P(F \in X) = P(\omega : F(\omega) \in X)$.

Rademacher Functions: (Figures 2 and 3)

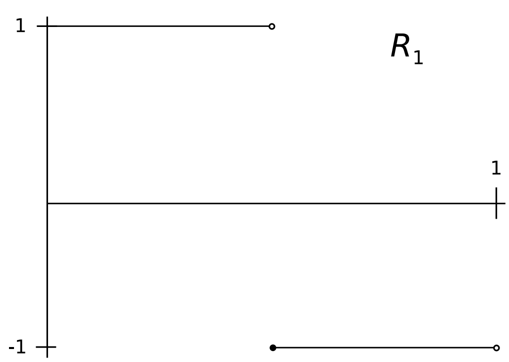

Figure 2

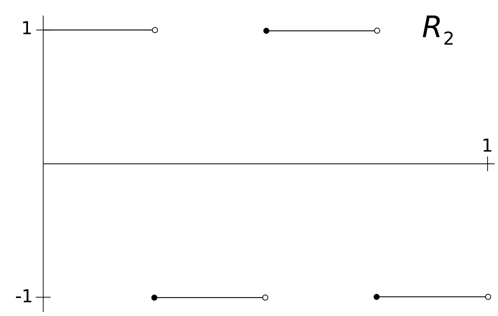

Figure 3

For $n = 1, 2, \ldots,$

$$r_n(\omega) = \sum_{k=1}^{2^m} (-1)^{k+1} 1_{[\frac{k-1}{2^n}, \frac{k}{2^n})}(w).$$

Note that $P(r_n = 1) = \frac{1}{2} = P(r_n = -1)$. Let 1 represent heads and -1 represent tails. We now have a model for flipping coins. $\mathbb{E}[r_n] = 0$.

Independence:

Formally, independence means that $P(r_{n_1} = e_1 \mid r_{n_2} = e_2) = P(r_{n_1} = e_1)$. That is, the probability that $r_{n_1} = e_1$ when we know that $r_{n_2} = e_2$ is the same as the probability that $r_{n_1} = e_1$ in the absence of additional information. We may assume that e_1 and e_2 are elements of $\{-1, 1\}$. (Events outside this space have 0 probability automatically.) With this condition, we see that

$$P(r_{n_1} = e_1, r_{n_2} = e_2, \ldots, r_{n_k} = e_k) = \prod_{i=1}^{k} P(r_{n_i} = e_i).$$

We use commas here to denote the probability of several events happening simultaneously.

$$\mathbb{E}[\prod_{i=1}^{k} r_{n_i}] = \prod_{i=1}^{k} \mathbb{E}[r_{n_i}] = 0.$$

$$\mathbb{E}[r_i r_j] = \delta_{i,j} \text{ (orthonormal)}.$$

$$R_n(w) = \sum_{i=1}^{n} r_i(w) \text{ (random walks)}.$$

Theorem 1.

$$\lim_{n\to\infty} \frac{R_n}{n} = 0 \textit{ (almost surely).}$$

Proof. $\mathbb{E}[(R_n)^4] = n + 3n(n-1)$. Therefore,

$$\sum_{n=1}^{\infty} \mathbb{E}\Big[\Big(\frac{R_n}{n}\Big)^4\Big] < \infty,$$

which implies

$$\mathbb{E}\Big[\sum_{n=1}^{\infty}\Big(\frac{R_n}{n}\Big)^4\Big] < \infty.$$

Hence

$$\sum_{n=1}^{\infty}\left(\frac{R_n}{n}\right)^4 < \infty \text{ (almost surely)},$$

and therefore

$$\lim_{n\to\infty} \frac{R_n}{n} = 0 \text{ (almost surely)}.$$

□

Now instead of having our random variables take values in $\{-1, 1\}$ we would like them to take values in $\{0, 1\}$ which is more useful for counting the number of heads occurs when you flip a coin. So the following construction toward that end is due to Steinhaus. Renormalize so that the range of our random variables is in $\{0, 1\}$

$$X_k(w) = \frac{1 - r_k(w)}{2},$$

and then take the n^{th} partial sum

$$S_n(w) = \sum_{k=1}^{n} X_k(w).$$

Now we see that

$$\lim_{n\to\infty} \frac{S_n}{n} = \frac{1}{2} \text{ (almost surely)},$$

and we have a random variable that takes values in $\{0,1\}$ and which has expectation $\frac{1}{2}$. Suppose that $w \in [0,1)$. Then, $x_n(w)$ gives the n^{th} digit of the binary expansion of w. It follows that

$$w = \sum_{n=1}^{\infty} \frac{X_n(w)}{2^n}.$$

Davar Khoshnevisan

Let $\{x_n\}$ be sequence of indepedent random variables with $x_n = 0$ or 1 with probability $\frac{1}{2}$. We can do the same thing for a larger set of possible outcomes, i.e. let $\{x_n\}$ be a sequence of random integers chosen from the set $\{0,1,2,\ldots,k\}$ with probability $\frac{1}{k}$ for each element of the set. Now, suppose that we choose a random real number from the interval $[0,1]$. What is the probability that this random number lies in $[0,1/3]$? It will be one third. Once we know that our number lies in $[0,1/3]$, we can subdivide this interval in three parts. What are the odds that our number lies in $[1/9,2/9]$? The probability is again one third. This suggests that we choose a random integer sequence as described before with $k=3$. Then,

$$X = \sum_{n=1}^{\infty} \frac{x_n}{3^n}$$

gives us a random number in $[0,1]$. Let $F(x) = P\{X < x\}$. Then we have the graph shown in Figure 4.

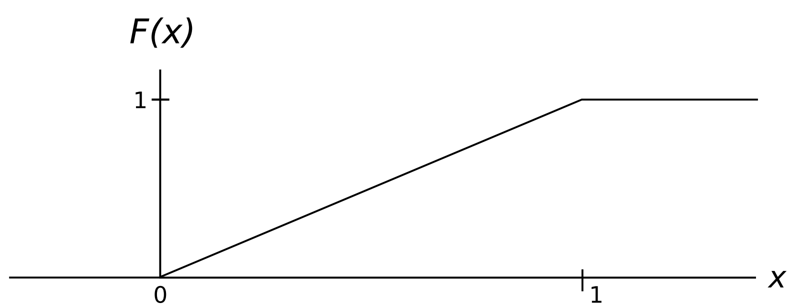

Figure 4

We can also play this game on Cantor sets. Recall the construction of the Cantor set: Begin with the interval $[0,1]$. Remove the open middle third. We are now left with two intervals and we remove the open middle third of each of these. Repeat this *ad infinitum*. Figure 5 illustrates the process.

Figure 5

Note that elements of the Cantor set have ternary expansions containing only 0's and 2's. In addition, any number in $[0, 1]$ with such a ternary expansion is in the Cantor set. Cantor observed that elements of the Cantor set are in one to one correspondence with elements of $[0, 1]$ by converting 0 and 2 in a ternary expansion to 0 and 1 in a binary expansion of a number in $[0, 1]$. How do we sample uniformly from the Cantor set? Choose a random sequence consisting of elements of $\{0, 2\}$. Zero and two will have equal probability, so we can construct a graph of $F(x) = P\{X \leq x\}$, where X is chosen randomly from the Cantor set. This is shown in Figure 6.

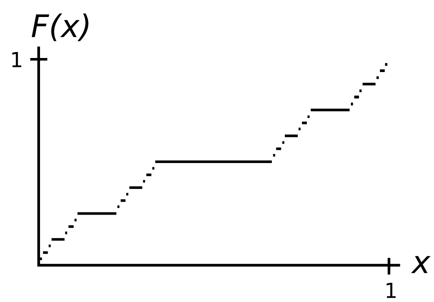

Figure 6

In the fractal world, this is often known as the Devil's Staircase, although it has a long history in probability. This function is continuous and has derivative 0 almost everywhere. However,

$$\int_0^1 F'(x)dx = 0$$

while $F(1) - F(0) = 1$. This is an example of where the fundamental theorem of calculus fails once we omit some hypothesis. In particular, $F(x)$ is not differentiable at every point in $[0, 1]$, but even more importantly $F(x)$ is not absolutely continuous with respect to Lesbesgue measure.

Lattice Paths:
Choose some $n \in \mathbb{N}$. Start at the origin and move to the right a distance $1/n$. Simultaneously, move either up or down a distance $1/n$. Continue this process until you reach x-coordinate 1 (Figure 9). This is sometimes referred to as a lattice path. Essentially, we are random walking with time and distance step $1/n$.

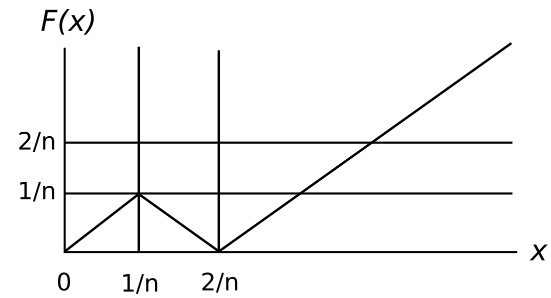

Figure 7

Scalar multiples of these paths form a dense subset of $C[0,1]$. We can study average properties for the family of paths with a given step size by taking up and down movements to be equally probable. For instance, on average (root mean square), the family of paths with step size $1/n$ ends a distance $\sqrt{n}$ away from the x-axis. An average step then moves us a distance $1/\sqrt{n}$ away from the origin. Lattice paths give us discrete Brownian motion. It is not surprising then that the set of all one-dimensional Brownian motions on $[0,1]$ is $C[0,1]$.

Bichromatic Coloring and Ramsey Numbers: Consider the complete graph on N vertices, i.e. the graph on N vertices where a unique edge connects each pair of vertices. See figure 8.

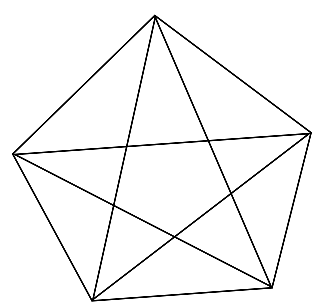
Figure 8: Complete Graph on 5 Vertices

We'll denote the complete graph on N vertices by K_N. The *Ramsey number* R_n is the smallest integer N such that any coloring of the edges of K_N yields a monochromatic subgraph on n vertices. The best known bounds on R_n are given by

$$c_1 n 2^{n/2} \leq R_n \leq C_2 2^n n^{c_3}.$$

The left inequality is proven by probability and the right by combinatorics. We can prove the left one without too much trouble: given a graph K_N, randomly color each edge either red or blue with equal probability. A subgraph K_n has $\binom{n}{2}$ edges, so the probability that this subgraph is monochromatically colored is $2/2^{\binom{n}{2}} = 2^{1-\binom{n}{2}}$. The number of n-vertex subgraphs of K_N is $\binom{N}{n}$, so the probability that K_N contains an n-vertex monochromatic subgraph is less than or equal to $\binom{N}{n}2^{1-\binom{n}{2}}$. (Here, we've used the Boolean inequality: $P(A \cup B) \leq P(A) + P(B)$.) Now, if we choose N such that this probability is less than 1 then $R_n > N$. Rounding down if necessary, we can use $2^{n/2}$.

It was conjectured by Paul Erdös that $\lim_{n\to\infty} \frac{\ln R_n}{n}$ exists.