Chapter 1

Universal examples

1.1 The Bernoulli process

First description: Bernoulli random variables Y_i for $i = 1, 2, 3, \ldots$ independent with $P[Y_i = 1] = p$ and $P[Y_i = 0] = 1 - p$.

Second description: Binomial random variables X_n for $n = 0, 1, 2, 3, \ldots$ with $X_0 = 0$ and $X_n = Y_1 + \cdots + Y_n$. Then

$$P[X_n = k] = \binom{n}{k} p^k (1-p)^{n-k}. \tag{1.1}$$

Third description: Waiting time random variables T_j for $j = 0, 1, 2, 3, \ldots$. Thus T_0 is the first i with $Y_i = 1$ and for $j \geq 1$ the random variable T_j is the difference between the $j+1$ value of i with $Y_i = 1$ and the jth value of i with $Y_i = 1$. The conditional distribution is

$$P[T_j = k \mid T_s, s < j] = (1-p)^{k-1} p. \tag{1.2}$$

It follows that the T_j are independent and

$$P[T_j = k] = (1-p)^{k-1} p. \tag{1.3}$$

So these are shifted geometric random variables

Fourth description: Renewal time random variables A_r for $r = 1, 2, 3, \ldots$ defined by $A_r = \min\{n \mid X_n = r\}$. We have

$$P[A_r = m] = P[X_{m-1} = r-1] P[Y_r = 1] = \binom{m-1}{r-1} p^{r-1} (1-p)^{m-r} p = \binom{m-1}{r-1} p^r (1-p)^{m-r}. \tag{1.4}$$

Note that

$$A_r = T_0 + \cdots + T_{r-1}, \tag{1.5}$$

so this gives the distribution for a sum of shifted geometric random variables.

Let us compute

$$P[A_1 = m_1, \ldots, A_k = m_k \mid X_n = k] = \frac{P[A_1 = m_1, \ldots, A_k = m_k, X_n = k]}{P[\mid X_n = k]} \tag{1.6}$$

This is

$$= \frac{p(1-p)^{m_1-1}p(1-p)^{m_2-m_1-1}\cdots p(1-p)^{m_k-m_{k-1}-1}(1-p)^{n-m_k}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{1}{\binom{n}{k}}. \tag{1.7}$$

This says given a total number of events, the occurrences of the set of times at which they take place is uniform. Another way to think of it is to consider the $(n)_k = n(n-1)(n-2)\cdots(n-k+1)$ injective functions from a k element set to an n element set. These each give a way of choosing the points in order. The corresponding number of subsets is $\binom{n}{k} = (n)_k/k!$. So the uniform probability is given by $1/\binom{n}{k} = k!/(n)_k$.

It is fairly obvious that the increments of the X_n are independent. However here is a verification of this fact that uses the conditional uniformity in a way that will turn out to be illuminating. The task is to compute the joint distribution of the increments of the X_n. Set $k = k_1 + \cdots + k_m$ and compute

$$P[X_{n_i} - X_{n_{i-1}} = k_i, i = 1, \ldots, m] = P[X_{n_i} - X_{n_{i-1}} = k_i, i = 1, \ldots, m \mid X_{n_m} = k]P[X_{n_m} = k] \tag{1.8}$$

This is

$$\frac{\prod_{i=1}^m \binom{n_i - n_{i-1}}{k_i}}{\binom{n_m}{k}} \binom{n_m}{k} p^k (1-p)^{n_m - k} = \prod_{i=1}^m \binom{n_i - n_{i-1}}{k_i} p^k (1-p)^{n_m - k}. \tag{1.9}$$

In turn this is

$$\prod_{i=1}^m P[X_{n_i} - X_{n_{i-1}} = k_i] = \prod_{i=1}^m [\binom{n_i - n_{i-1}}{k_i} p^{k_i}(1-p)^{n_i - n_{i-1} - k_i}]. \tag{1.10}$$

In other words,

$$\prod_{i=1}^m P[X_{n_i} - X_{n_{i-1}} = k_i] = \prod_{i=1}^m P[X_{n_i} - X_{n_{i-1}} = k_i]. \tag{1.11}$$

1.2 The Poisson process

Third description: Let $T_0, T_1, T_2, \ldots$ be independent exponential random variables with intensity parameter $\lambda > 0$. These are the waiting times. Thus

$$E[f(T_j)] = \int_0^\infty f(t)e^{-\lambda t}\lambda\, dt. \tag{1.12}$$

In particular, $P[T_j > t] = \exp(-\lambda t)$ for $t \geq 0$. Notice that $\lambda e^{-\lambda t}$ is a probability density measured in inverse seconds, while $e^{-\lambda t}$ is a probability.

The motivation for this definition comes from the identity

$$P[T_j > s + t \mid T_j > s] = P[T_j > t]. \tag{1.13}$$

This says that if one has waited for a time $s \geq 0$ and nothing has happened, the the chances for waiting an additional time $t \geq 0$ are just as if one were starting again at the beginning. In other words, there is no memory of the original time s.

Fourth description: Let $A_r = T_0 + \cdots T_{r-1}$. These are the renewal times. This is the sum of independent random variables, so the density is obtained by convolution. It is easy to see by induction that the density is given by the Gamma density, so

$$E[f(A_r)] = \int_0^\infty f(t) \frac{(\lambda t)^{r-1}}{(r-1)!} e^{-\lambda t} \lambda \, dt. \tag{1.14}$$

In fact, we can easily compute the convolution

$$\int_0^t \lambda e^{-\lambda(t-s)} \frac{(\lambda s)^{r-1}}{(r-1)!} e^{-\lambda(t-s)} \lambda \, ds = \frac{(\lambda t)^r}{(r-1)!} e^{-\lambda t} \lambda \int_0^t s^{r-1} \, ds = \frac{(\lambda t)^r}{(r)!} e^{-\lambda t} \lambda. \tag{1.15}$$

Notice that $\lambda \frac{(\lambda t)^{r-1}}{(r-1)!} e^{-\lambda t}$ is a probability density measured in inverse seconds.

Second description: Define X_t to be the largest n with $A_n \leq t$. Thus $X_t < k+1$ if and only if $A_{k+1} > t$. From this we see that

$$P[X_t = k] = P[A_{k+1} > t] - P[A_k > t]. \tag{1.16}$$

On the other hand, we use integration by parts to compute

$$P[A_{k+1} > t] = \int_t^\infty \frac{(\lambda s)^k}{k!} e^{-\lambda s} \lambda \, ds = P[A_k > t] + \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \tag{1.17}$$

We conclude that

$$P[X_t = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \tag{1.18}$$

Thus X_t is Poisson with mean λt.

Now we show that the conditional distribution of $A_1, \ldots, A_k$ given $X_t = k$ is the distribution of uniform order statistics. That is, the conditional density of $A_1, \ldots, A_k$ given $X_t = k$ is

$$p(a_1, a_2, \ldots, a_k) = \frac{k!}{t^k} \tag{1.19}$$

for $0 \leq a_1 < a_2 < \cdots < a_k \leq t$. Notice that this is probability density function with dimension inverse seconds to the k power. It is what results from taking k independent uniform random variables on $[0, t]$ and arranging them in order.

Since the joint density of $T_0, \ldots, T_k$ is $\lambda^{k+1} e^{-\lambda(t_0 + \cdots + t_k)}$, it follows that the joint density of $A_1, \ldots A_{k+1}$ is $\lambda^{k+1} e^{-\lambda a_{k+1}}$ for increasing $0 < a_1 < a_2 < \cdots <$

a_{k+1}. This is because the Jacobian of the transformation $a_i = t_0 + \cdots + t_{i-1}$ is equal to 1. Now if we compute the joint density of $A_1, \ldots, A_k$ given $X_t = k$, then we must integrate over the possible values of A_{k+1} greater than t. This gives the result

$$p(a_1, a_2, \ldots, a_k) = \frac{\int_t^\infty \lambda^{k+1} e^{-\lambda a_{k+1}}\, da_{k+1}}{P[X_t = k]} = \frac{\lambda^k e^{-\lambda t}}{\frac{(\lambda t)^k}{k!} e^{-\lambda t}} \tag{1.20}$$

for $a_1 < a_2 < \cdots < a_k \leq t$. This simplifies to give the result above.

We conclude by proving the independent increment property of the Poisson process:

$$\prod_{i=1}^m P[X_{t_i} - X_{t_{i-1}} = k_i] = \prod_{i=1}^m P[X_{n_i} - X_{n_{i-1}} = k_i]. \tag{1.21}$$

The proof uses the previous result on the conditional distribution of the renewal times. Set $k = k_1 + \cdots + k_m$ and compute

$$P[X_{t_i} - X_{t_{i-1}} = k_i, i = 1, \ldots, m] = P[X_{t_i} - X_{t_{i-1}} = k_i, i = 1, \ldots, m \mid X_{t_m} = k] P[X_{t_m} = k] \tag{1.22}$$

The conditional probability is the probability that of k uniformly chosen points in $[0, t_m]$ we have k_i in $[t_{i-1}, t_i]$ for each $i = 1, \ldots, m$. Thus we have

$$P[X_{t_i} - X_{t_{i-1}} = k_i, i = 1, \ldots, m \mid X_{t_m} = k] = \frac{k!}{k_1! \cdots k_m!} \prod_{i=1}^m \left(\frac{t_i - t_{i-1}}{t_m} \right)^{k_i}. \tag{1.23}$$

Thus

$$P[X_{t_i} - X_{t_{i-1}} = k_i, i = 1, \ldots, m] = \prod_{i=1}^m \frac{(\lambda(t_i - t_{i-1})^{k_i}}{k_i!} e^{-\lambda(t_i - t_{i-1})}. \tag{1.24}$$

1.3 Simple random walk

We have $Z_1, Z_2, Z_3, \ldots$ independent with $P[Z_i = 1] = p$ and $P[Z_i = -1] = q$, where $p + q = 1$. The simple random walk is defined by $X_0 = x$ and $X_n = x + Z_1 + \cdots + Z_n$. We denote the corresponding probabilities by P_x.

If we start at zero, we get

$$P_0[X_n = y] = \binom{n}{\frac{n+y}{2}} p^{\frac{n+y}{2}} q^{\frac{n-y}{2}} \tag{1.25}$$

when y has the same parity as n. This is because $(n + y)/2$ minus $(n - y)/2$ is y. Note that if we start at some other x, then $P_x[X_n = y]$ is obtained from the above formula by replacing y by $y - x$.

Let T_y be the first time $n \geq 0$ that $X_n = y$. Let $c < d$. Start the process at x with $c \leq x \leq d$. We want to compute

$$\phi(x) = P_x[T_d < T_c]. \tag{1.26}$$

This is the probability that one achieves d before sinking to c.

The trick is to compute that

$$\phi(x) = p\phi(x+1) + q\phi(x-1). \tag{1.27}$$

As we shall see, this equation is a consequence of the Markov property of the random walk. It is called a backward equation, because the variable is the initial condition back at the beginning of the walk.

When $p \neq q$ the general solution of this equation is

$$\phi(x) = A + B(q/p)^x. \tag{1.28}$$

The boundary conditions are $\phi(c) = 0$ and $\phi(d) = 1$. From the first condition one sees that $A + B(q/p)^c = 0$. The second boundary condition may then be used to solve for the probabilities.

Next we compute $P_x[T_d < \infty]$. Notice that as c decreases, the events $[T_d < T_c]$ increase to the event $[T_d < \infty$. So all we have to do is to take the limit of the probabilities as $c \to -\infty$. It is clear that if $p > q$ then the limit is A, while if $p < q$ the limit is $B(q/p)^x$. So for $p > q$ the answer is 1, while for $p < q$ the answer is $(q/p)^{x-d} = (p/q)^{d-x}$.

When $p \neq q$ the random walk is transient. This means that it returns to x only finitely many times. One way to see this is to consider the case $p > q$ and use the above result. It is easy to compute that the probability of a return to x is $2q < 1$. Each time it returns to x, the process faces a task of making yet one more such return. The probability of k returns is $(2q)^k$. So the probability of infinitely many returns is zero. As we shall see, this argument uses the strong Markov property.

It remains to consider $p = q = 1/2$. In this case the general solution of the equation is

$$\phi(x) = A + Bx. \tag{1.29}$$

The first boundary condition gives $A + Bc = 0$. If we want to compute $P[T_d < \infty]$, then we can can take $c \to -\infty$ as before. The only way for this to give a reasonable result is to have the limit $A = 1$.

When $p = q = 1/2$ the random walk is recurrent. This means that it returns to x infinitely many times. The reason for this is simple. The probability of a return is one. After a return, the process starts fresh, and again the probability of a return is one. In this way, we see that the probability of k returns is one. It follows that the probability of infinitely many returns is zero.

1.4 The strong Markov property

In the following we shall often consider a σ-algebra of events as incorporating limited information about an experiment. Often the σ-algebra is generated by the values of functions $X_1, \ldots, X_n$. This just means that it is generated by the events $[X_i \in B]$. It is an important theorem that if a random variable Y is measurable with respect to the σ-algebra generated by $X_1, \ldots, X_n$, then Y is a function $Y = g(X_1, \ldots, X_n)$ of these random variables.

Let $\mathcal{F}$ be a σ-algebra of events. If A is an event, then the conditional probability of A with respect to $\mathcal{F}$ is a random variable $P[A \mid \mathcal{F}]$. It is characterized by the following two properties.

1. $P[A \mid \mathcal{F}]$ is measurable with respect to $\mathcal{F}$.
2. If C is in $\mathcal{F}$, then $P[A, C] = E[P[A \mid \mathcal{F}]1_C]$.

Let $\mathcal{F}$ be a σ-algebra of events. If Y is a random variable, then the conditional expectation of Y with respect to $\mathcal{F}$ is another random variable $E[Y \mid \mathcal{F}]$. It is characterized by the following two properties.

1. $E[Y \mid \mathcal{F}]$ is measurable with respect to $\mathcal{F}$.
2. If C is in $\mathcal{F}$, then $E[Y 1_C] = E[E[Y \mid \mathcal{F}]1_C]$.

Let $\mathcal{F}_m = \sigma(X_0, X_1, \ldots X_m)$ be the σ-algebra generated by the first steps $X_0, X_1, ..., X_m$. of the random walk. The functions measurable with respect to $\mathcal{F}_m$ are just the functions of $X_0, \ldots, X_m$. For each m let X_m^+ be the shifted random walk $X_m, X_{m+1}, X_{m+2}, \ldots$. The Markov property says that

$$P[X_m^+ \in B \mid \mathcal{F}_m] = P_{X_m}[X \in B]. \tag{1.30}$$

Here B is an arbitrarily measurable set of paths. In other words, the conditional probability for the future of m, which a priori could depend on $X_0, X_1, \ldots, X_m$, actually only depends on the present X_m. Furthermore, the probability is just the same as if one started fresh at the point X_m.

The Markov property could also be expressed in expectations by

$$E[f(X_m^+) \mid \mathcal{F}_m] = E_{X_m}[f(X)]. \tag{1.31}$$

Here f is an arbitrary bounded measurable function on the space of paths.

Here is an example of the Markov property at work. Consider the symmetric simple random walk. Then $P[X_5 = 1 \mid \mathcal{F}_2] = P_{X_2}[X_3 = 1] = \frac{1}{8}1_{X_2=-2} + \frac{3}{8}1_{X_2=0} + \frac{3}{8}1_{X_3=2}$. Hence $P[X_5 = 1] = E[P[X_5 = 1 \mid \mathcal{F}_2]] = \frac{1}{8}\frac{1}{4} + \frac{3}{8}\frac{1}{2} + \frac{3}{8}\frac{1}{4} = \frac{5}{16}$. This would not be difficult to compute directly as the binomial probability $10\frac{1}{2^5}$.

Consider a random variable τ with values in $0, 1, 2, 3, \ldots$ or ∞. It is called a stopping time if for each n the event $[\tau \leq n]$ is in $\mathcal{F}_n$. In other words, it can be determined whether $\tau \leq n$ by knowing $X_0, \ldots, X_n$. One does not have to look into the future.

It is equivalent to require that for each n the event $[\tau = n]$ is in $\mathcal{F}_n$. For in that case $[\tau \leq n]$ is the union of $[\tau = k]$ for $k = 0, 1, 2, \ldots, n$, and each such $[\tau = k]$ is in $\mathcal{F}_k$ and hence in $\mathcal{F}_n$. On the other hand, we can define $[\tau = n]$ as $[\tau \leq n]$ and not $[\tau \leq n - 1]$.

We can also define a new σ-algebra $\mathcal{F}_\tau$ defined by $C \in \mathcal{F}_\tau$ if and only if for each n we have $C \cap [\tau = n] \in \mathcal{F}_n$. It can be shown that a random variable measurable with respect to this σ-algebra is a function of the random variables $X_0, X_1, X_2, X_3, \ldots, X_\tau, X_\tau, X_\tau, \ldots$.

In the following we shall need a new path $\emptyset$ that is defined no times at all. It may be regarded as starting at a non-existent point X_∞. If we try to start a path X at the non-existent point X_∞, then it is regarded as the path $\emptyset$.

Let τ be a stopping time. Define a new process X_τ^+ to be X_m^+ on $\tau = m$ and to be $\emptyset$ on $\tau = \infty$. The strong Markov process says that if B is a subset of the set of all paths (augmented by the empty path), then

$$P[X_\tau^+ \in B \mid \mathcal{F}_\tau] = P_{X_\tau}[X \in B]. \tag{1.32}$$

In terms of expectations, this says

$$E[f(X_\tau^+) \mid \mathcal{F}_\tau] = E_{X_\tau}[f(X)]. \tag{1.33}$$

Here is the proof. Let C be an event in $\mathcal{F}_\tau$. We compute

$$E[E[f(X_\tau^+)1_C] = \sum_{m=0}^{\infty} E[f(X_m^+)1_{C\cap[\tau=m]} + E[f(\emptyset)1_{C\cap[\tau=\infty]}]. \tag{1.34}$$

By the ordinary Markov property

$$E[f(X_m^+)1_{C\cap[\tau=m]}] = E[E_{X_m}[f(X)]1_{C\cap[\tau=m]}] \tag{1.35}$$

this is

$$\sum_{m=0}^{\infty} E[E_{X_m}[f(X)]1_{C\cap[\tau=m]} + E[E_{X_\infty}[f(X)]1_{C\cap[\tau=\infty]}] = E[E_{X_\tau}[f(X)]1_C]. \tag{1.36}$$

Here is an example of the strong Markov property at work. Let τ be the hitting time for the set $\{-1, 3\}$. Then $P[X_{\tau+3} = 0 \mid \mathcal{F}_\tau] = P_{X_\tau}[X_3 = 0] = \frac{1}{8}1_{X_\tau=3} + \frac{3}{8}1_{X_\tau=-1}$. We may conclude that $P[X_{\tau+3} = 0] = \frac{1}{8}\frac{1}{4} + \frac{3}{8}\frac{3}{4} = \frac{5}{16}$.

We have used the strong Markov property in the reasoning about transience and recurrence of random walk. Thus if τ^1 and τ^2 are the time of the first and second returns to x, we have by the strong Markov property

$$P_x[\tau^2 < \infty \mid \mathcal{F}_\tau^1] = P_{X_\tau^1}[\tau^1 < \infty] = P_x[\tau^1 < \infty]1_{\tau^1<\infty}. \tag{1.37}$$

Hence

$$P[\tau^2 < \infty] = E[P[\tau^2 < \infty \mid \mathcal{F}_\tau^1]] = (P_x[\tau^1 < \infty])^2. \tag{1.38}$$

1.5 Martingales

A stochastic process $M_n = \phi(X_n)$ for $n = 0, 1, 2, 3, \ldots$ is a martingale if

$$E[M_{n+1} \mid \mathcal{F}_n] = M_n. \tag{1.39}$$

For the random walk with $p = q = 1/2$ we see that $M_n = X_n$ is a martingale. For the random walk with $p \neq q$ we see that $M_n = (q/p)^{X_n}$ is a martingale.

It is not hard to see that if M_n is a martingale, then

$$E[M_n] = E[M_0]. \tag{1.40}$$

It may be shown that if τ is a stopping time, then $M_{n\wedge\tau}$ is a martingale. It follows that for each n we also have

$$E[M_{n\wedge\tau}] = E[M_0]. \tag{1.41}$$

Let us apply this to the random walk with $p = q = 1/2$. Let $c < x < d$ and let τ be the first time n that X_n is either c or d. Then

$$E[X_{n\wedge\tau}] = x. \tag{1.42}$$

By the dominated convergence theorem

$$E[X_\tau] = x. \tag{1.43}$$

This says that

$$cP[X_\tau = c] + dP[X_\tau = d] = x. \tag{1.44}$$

This may be solved to get the hitting probabilities.

Let us apply this to the random walk with $p \neq q$. Let $c < x < d$ and let τ be the first time n that X_n is either c or d. Then

$$E[(q/p)^{X_{n\wedge\tau}}] = (q/p)^x. \tag{1.45}$$

By the dominated convergence theorem.

$$E[(q/p)^{X_\tau}] = (q/p)^x. \tag{1.46}$$

This says that

$$(q/p)^c P[X_\tau = c] + (q/p)^d P[X_\tau = d] = (q/p)^x. \tag{1.47}$$

This may be solved to get the hitting probabilities.

Now suppose $p > q$ and take $c < x$ and define τ as the first time n that $X_n = c$. Then

$$E[(q/p)^{X_{n\wedge\tau}}] = (q/p)^x. \tag{1.48}$$

Notice that $(q/p)^{X_n\wedge\tau} \leq (q/p)^c$. By the law of large numbers $X_n \to \infty$ as $n \to \infty$. So the only possible limits for $(q/p)^{X_n\wedge\tau}$ are 0 and $(q/p)^c$. By the dominated convergence theorem.

$$E[(q/p)^{X_\tau}] = (q/p)^x. \tag{1.49}$$

This says that

$$(q/p)^c P[\tau M\infty] = (q/p)^x. \tag{1.50}$$

This may be solved to get the hitting probability $P[\tau < \infty] = (q/p)^{x-c}$.

The use of martingales to compute these hitting probabilities is an example of a forward method, since the initial condition is fixed and one follows the probability forward in time.

1.6 The Wiener process

The Wiener process is also called the Einstein model of Brownian motion.

Let $L^2[0,1]$ be the Hilbert space of real functions f with

$$\|f\|^2 = \int_0^1 f(t)^2\,dt < \infty. \tag{1.51}$$

The inner product is

$$\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt. \tag{1.52}$$

Let ϕ_n be an orthonormal basis for $L^2[0,1]$. Let $Z_1, Z_2, Z_3, \ldots$ be independent standard Gaussian (normal) random variables. (Standard means mean zero and variance one.) If

$$f = \sum_n c_n \phi_n \tag{1.53}$$

define the random variable

$$A(f) = \sum_n c_n Z_n \tag{1.54}$$

Then $A(f)$ is Gaussian with mean zero and covariance

$$E[A(f)A(g)] = \langle f, g\rangle. \tag{1.55}$$

This stochastic process is called integrated white noise.

Intuition: The intuitive interpretation is that

$$A(f) = \int_0^1 a(t)f(t)\,dt, \tag{1.56}$$

where $a(t)$ is white noise. Here

$$a(t) = \sum_n Z_n \phi_n(t). \tag{1.57}$$

Since all basis vectors contribute roughly the same amount, this is the justification for the terminology white noise. (All colors contribute equally to white light.)

Of course the random function $a(t)$ does not exist. However, it is often convenient to compute as if it did exist. For instance, we have

$$\sum_n \phi_n(s)\phi_n(t) = \delta(t-s), \tag{1.58}$$

and so

$$E[a(s)a(t)] = \delta(t-s), \tag{1.59}$$

where the delta function is also a function that does not exist.

For $0 \leq t \leq 1$, let $B_t = A(1_{[0,t]})$. This is called the standard Brownian motion (on the unit interval). These random variables are Gaussian with mean zero and variance t. Their covariance are

$$E[B_s B_t] = \min(s,t). \tag{1.60}$$

Intuition: We have

$$B_t = \int_0^t a(t')\, dt' \tag{1.61}$$

so that Brownian motion is the integral of the (non-existent) white noise.

A standard Brownian motion B_t for all $t \geq 0$ starting with $B_0 = 0$ may be constructed by piecing together independent copies B_t^n in such a way that $B_t = B_n + B_{t-n}^{n+1}$ on the interval $[n, n+1]$. Again we have the same formulas for the mean and covariance.

We can construct a more general Brownian motion process

$$X_t = x + \sigma B_t + \mu t. \tag{1.62}$$

This has mean $x + \mu t$ and variance $\sigma^2 t$. It is the Brownian motion with starting point x and drift coefficient μ and diffusion parameter σ. (Note that $\sigma^2/2$ is sometimes called the diffusion constant.)

For convenience we work with standard Brownian motion B_t. It is Gaussian with independent increments. Now we show that there is a version that has continuous paths.

To do this, we make a choice of the orthonormal basis. We take it to consist of the Haar wavelet functions $\phi_{n,k}$ defined for $n = -1, 0, 1, 2, 3, \ldots$ and $0 \leq k \leq 2^n - 1$. The case $n = -1$ is exceptional; there is no index k, and the function $\phi_{-1} = 1$. For all other cases the function $\phi_{n,k}$ is obtained from the mother wavelet function $\phi(t)$ which is 1 on $(0, \frac{1}{2})$ and -1 on $(\frac{1}{2}, 1)$ and zero elsewhere on the line. The functions $\phi_{n,0}(t)$ are defined by shrinking $\phi_{n,0}(t) = 2^{\frac{n}{2}} \phi(2^n t)$. The functions $\phi_{n,k}$ are defined by translating $\phi_{n,k}(t) = \phi_{n,0}(t - \frac{k}{2^n})$.

Intuition: The white noise is

$$a(t) = \sum_n \sum_k Z_{n,k} \phi_{n,k}(t). \tag{1.63}$$

Let

$$S_{n,k}(t) = \int_0^t \phi_{n,k}(t')\, dt'. \tag{1.64}$$

The function $S_{-1} = 1$. For all other cases the function $S_{n,k}$ is obtained from the integrated mother wavelet function $S(t)$ which is t on $(0, \frac{1}{2})$ and $1 - t$ on $(\frac{1}{2}, 1)$ and zero elsewhere on the line. The functions $S_{n,0}(t)$ are given by $S_{n,0}(t) = 2^{-\frac{n}{2}} S(2^n t)$. The functions $S_{n,k}$ are given by translating $S_{n,k}(t) = S_{n,0}(t - \frac{k}{2^n})$. The key fact is that for fixed $n \geq 0$ the 2^n functions $S_{n,k}(t)$ are bounded by $2^{-\frac{n}{2}} \frac{1}{2}$ and have disjoint supports.

We define the canonical version of Brownian motion by

$$B_t = \sum_{n,k} Z_{n,k} S_{n,k}(t). \tag{1.65}$$

We will show that with probability one the convergence is uniform on the unit interval. This shows that there is a version of Brownian motion with continuous paths.

The Gaussian random variables are not bounded, but it is easy to get a good estimate on the probability of a large value. In fact, for $c \geq 1$ the probability that $|Z| \geq c$ is bounded by a constant times $e^{-\frac{c^2}{2}}$.

Let $a > 1$. Then the probability that $\max_k |Z_{n,k}| \geq a^n$ is bounded by a constant times $2^n e^{-\frac{a^{2n}}{2}}$. It is thus easy to prove that

$$\sum_n P[\max_k |Z_{n,k}| \geq a^n] < \infty. \tag{1.66}$$

It follows from the first Borel-Cantelli lemma that with probability one $\max_k |Z_{n,k}| \leq a^n$ eventually.

The trick is to use this to estimate the maximum value

$$\max_{0 \leq t \leq 1} \sum_k |Z_{n,k} S_{n,k}(t)| \leq \frac{1}{2} 2^{-\frac{n}{2}} \max_k |Z_{n,k}| \leq \frac{1}{2} 2^{-\frac{n}{2}} a^n. \tag{1.67}$$

If we take $a < 2^{\frac{1}{2}}$ then this is an exponential bound that holds eventually with probability one. It follows that the series defining the Brownian motion converges uniformly with probability one.