Random Walk

- Let Z_i be iid with $Pr[Z_i = +s] = Pr[Z_i = -s] = 0.5$
- Define a continuous-time process $X(t)$ by

$$
\begin{aligned}
X(nT) &= Z_1 + Z_2 + \cdots + Z_n \\
X(t) &= X(nT), \quad \text{for } \; nT \le t < (n+1)T
\end{aligned}
$$

- Viewed as a discrete-time process, $X(nT)$, is a Markov chain.
- Called a (one dimensional) random walk
- It is the position after n random steps
- We defined $X(t)$ by piece-wise constant interpolation of $X(nT)$
- We could have also use piece-wise linear interpolation

- We have $EZ_i = 0$ and $E[Z_i^2] = s^2$
- Hence, $E[X(nT)] = 0$ and $E[X^2(nT)] = ns^2$
- For large n, $\frac{X(nT)}{s\sqrt{n}}$ would be Gaussian

$$Pr\left[\frac{X(nT)}{s\sqrt{n}} \leq y\right] \approx \Phi(y)$$

  where Φ is distribution function of standard Normal
- For any t, $X(t)$ is $X(nT)$ for $n = [t/T]$.
  Large n would mean large t. Hence

$$Pr[X(t) \leq ms] = Pr\left[\frac{X(t)}{s\sqrt{n}} \leq \frac{ms}{s\sqrt{n}}\right] \approx \Phi\left(\frac{m}{\sqrt{n}}\right), \quad \text{for large } t$$

- We are interested in limit of this process as $T \to 0$

- Consider $t = nT$

$$E[X^2(t)] = ns^2 = s^2 \frac{t}{T}$$

- If we let $T \to 0$ then the variance goes to infinity (the process goes to infinity) unless we let s also to go to zero.
- We actuall need s^2 to go to zero at the same rate as T.
- So, we keep $s^2 = \alpha T$ and let T go to zero.
- Define

$$W(t) = \lim_{T \to 0, s^2 = \alpha T} X(t)$$

  This is called the Wiener Process or Brownian motion. This result is known as Donsker's theorem
- Let us intuitively see some properties of $W(t)$

- We have seen that for $n = [t/T]$,

$$Pr[X(t) \leq ms] \approx \Phi\left(\frac{m}{\sqrt{n}}\right)$$

- Let $w = ms$ and $t = nT$. Then

$$\frac{m}{\sqrt{n}} = \frac{w/s}{\sqrt{t/T}} = \frac{w}{\sqrt{t}}\sqrt{\frac{T}{s^2}} = \frac{w}{\sqrt{\alpha t}}$$

- $W(t)$ is limit of $X(t)$ as T goes to zero
- As T goes to zero, any t is 'large n'.
- Hence we can expect

$$Pr[W(t) \leq w] = \Phi\left(\frac{w}{\sqrt{\alpha t}}\right)$$

$$\Rightarrow \quad W(t) \sim \mathcal{N}(0,\ \alpha t)$$

- We had Z_i iid and defined

$$X(nT) = Z_1 + Z_2 + \cdots + Z_n$$

- Hence we get

$$X((m+n)T) - X(nT) = Z_{n+1} + \cdots + Z_{n+m}$$

  Thus, $X(nT)$ is independent of $X((m+n)T) - X(nT)$.
- Hence the $X(nT)$ process has independent increments
- Hence, we can expect $W(t)$ to be a process with independent increments

- $X((m+n+k)T) - X((n+k)T)$ and
  $X((m+n)T) - X(nT)$
  both are sums of m of the Z_i's
- Hence both would have the same distribution
- Thus $X(nT)$ would also have stationary increments.
- Hence we also expect $W(t)$ to have stationary increments
- Thus, $W(t)$ should be a process with stationary and independent increments and for each t, $W(t)$ is Gaussian with zero mean and variance proportional to t
- We will now formally define Brownian motion using these properties.

- Let $\{X(t),\ t \geq 0\}$ be a continuous-state continuous-time process

  This process is called a Brownian motion if

  1. $X(0) = 0$
  2. The process has stationary and independent increments
  3. For every $t > 0$, $X(t)$ is Gaussian with mean 0 and variance $\sigma^2 t$

- Let $B(t) = \frac{X(t)}{\sigma}$. Then, variance of $B(t)$ is t
- $\{B(t),\ t \geq 0\}$ is called standard Brownian Motion
- Let $Y(t) = X(t) + \mu$. Then $Y(t)$ has non-zero mean
- The mean can be a function of time
- $\{Y(t),\ t \geq 0\}$ is called Brownian motion with a drift

- Let $\{X(t),\ t \geq 0\}$ be a Brownian motion
- The process has stationary increments.
- Hence for $t_2 > t_1$, $X(t_2) - X(t_1)$ has the same distribution as $X(t_2 - t_1)$
- Thus, $X(t_2) - X(t_1)$ is Gaussian with zero mean and variance $\sigma^2(t_2 - t_1)$
- Since increments are also independent, we can show that all n^{th} order distributions are Gaussian

- We can calculate the autocorrelation function

$$
\begin{aligned}
R_X(t_1, t_2) &= E[X(t_1)X(t_2)] \\
&= E[X(t_1)\ (X(t_2) - X(t_1) + X(t_1))], \quad (\text{take } \ t_1 < t_2) \\
&= E[X(t_1)(X(t_2) - X(t_1))] + E[X^2(t_1)] \\
&= E[X(t_1)]\ E[X(t_2) - X(t_1)] + E[X^2(t_1)] \\
&= E[X^2(t_1)] \\
&= \sigma^2\ t_1
\end{aligned}
$$

- Since $E[X(t)] = 0, \ \forall t$, we have

$$\text{Cov}(X(t_1), X(t_2)) = E[X(t_1)X(t_2)] = \sigma^2 \ \min(t_1, t_2)$$

- Suppose we want the joint distribution of $X(t_1), X(t_2), \cdots, X(t_n)$
- Let $t_1 < t_2 < \cdots < t_n$
- Define random variables $Y_1, \cdots, Y_n$ by

$$Y_1 = X(t_1), \;\; Y_2 = X(t_2) - X(t_1), \;\; Y_3 = X(t_3) - X(t_2), \cdots$$

- We know Y_i are independent because the process has independent increments
- This transformation is invertible
- Hence we can get joint density of $X(t_1), \cdots X(t_n)$ in terms of joint density of $Y_1, \cdots, Y_n$
- This is how we can get n^{th} order density for any continuous-state process with independent increments

$$Y_1 = X(t_1), \;\; Y_i = X(t_i) - X(t_{i-1}), \; i = 2, \cdots, n$$

- The transformation is invertible

$$\begin{aligned} X(t_1) &= Y_1 \\ X(t_2) &= Y_1 + Y_2 \\ X(t_3) &= Y_1 + Y_2 + Y_3 \\ &\vdots \\ X(t_n) &= Y_1 + Y_2 + \cdots + Y_n \end{aligned}$$

- $Y_1, \cdots Y_n$ are independent and Gaussian and hence are Jointly Gaussian
- Hence $X(t_1), \cdots, X(t_n)$ are jointly Gaussian
- Thus all n^{th} order distributions are Gaussian

- $X(t_1), X(t_2), \cdots, X(t_n)$ are jointly Gaussian.
- We can write their joint density because we know the means, variances and covariances
- We can also write the density using the transformation considered earlier
  Let $t_1 < t_2 < \cdots < t_n$

$$f_X(x_1, \cdots, x_n; t_1, \cdots, t_n) \;=\; f_{Y_1}(x_1) f_{Y_2}(x_2 - x_1) \cdots f_{Y_n}(x_n - x_{n-1})$$

- Note that $Y_i = X(t_i) - X(t_{i-1})$ is Gaussian with mean zero and variance $\sigma^2(t_i - t_{i-1})$, $i = 1, \cdots, n$
  (Take $t_0 = 0$)

- Since all joint densities are Gaussian and are easy to write, we can also calculate conditional densities

$$
\begin{aligned}
f_{X(s)|X(t)}(x|b) &= \frac{f_{X(s)X(t)}(x,b)}{f_{X(t)}(b)} \quad (s<t) \\
&= \frac{f_{X(s)}(x)\ f_{X(t)-X(s)}(b-x)}{f_{X(t)}(b)} \\
&\propto e^{-\frac{x^2}{2s}}\ e^{-\frac{(b-x)^2}{2(t-s)}} \quad \text{(taking } \sigma^2 = 1\text{)} \\
&\propto \exp\left(-x^2\left(\frac{1}{2s}+\frac{1}{2(t-s)}\right)+\frac{bx}{t-s}\right) \\
&\propto \exp\left(-\frac{t}{2s(t-s)}\left(x^2-2\frac{sb}{t}x\right)\right) \\
&\propto \exp\left(-\frac{(x-bs/t)^2}{2s(t-s)/t}\right)
\end{aligned}
$$

- Hence the conditional density is Gaussian with mean bs/t and variance $s(t-s)/t$

- An important result is that Brownian motion paths are continuous
- Brownian motion is the limit of random walk where both s and T tend to zero
- Intuitively the paths should be continuous.
- The paths are continuous but non-differentiable everywhere
- This is a deep result

Hitting Times

- Let T_a denote the first time Brownian motion hits a. We take $a > 0$.

$$Pr[X(t) \geq a] = Pr[X(t) \geq a \mid T_a \leq t] \; Pr[T_a \leq t] + Pr[X(t) \geq a \mid T_a > t] \; Pr[T_a > t]$$

- Since Brownian motion paths are continuous, $Pr[X(t) \geq a \mid T_a > t] = 0$
- Brownian motion is a limit of symmetric random walk. Hence if we had already hit a sometime back, then now we are as likely to be above a as below it.

$$\Rightarrow \quad Pr[X(t) \geq a \mid T_a \leq t] = \frac{1}{2}$$

Thus

$$P[X(t) \geq a] = 0.5 Pr[T_a \leq t]$$

- Hence we get

$$\begin{aligned} Pr[T_a \leq t] &= 2\ Pr[X(t) \geq a] \\ &= \frac{2}{\sqrt{2\pi t}} \int_a^{\infty} e^{-\frac{x^2}{2t}}\ dx \\ &= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^{\infty} e^{-\frac{y^2}{2}}\ dy \end{aligned}$$

- Here we have assumed $a > 0$. For $a < 0$ the situation is similar. Hence the above is true even for $a < 0$ except that the lower limit becimes $|a|/\sqrt{t}$
- Another interesting consequence is the following

$$\begin{aligned} Pr[\max_{0 \leq s \leq t} X(s) \geq a] &= Pr[T_a \leq t], \text{ by continuity of paths} \\ &= 2Pr[X(t) \geq a] \end{aligned}$$

Geometric Brownian Motion

- Let $\{Y(t),\ t \geq 0\}$ is a Brownian motion with drift. Define
$$X(t) = e^{Y(t)}$$
- Then, $\{X(t),\ t \geq 0\}$ is called geometric Brownian motion. It is useful in mathematial finance
- Let $X_0, X_1, \cdots$ be time series of prices of a stock.
- Let $Y_n = X_n / X_{n-1}$ and assume Y_i are iid
$$X_n = Y_n X_{n-1} = Y_n Y_{n-1} X_{n-2} = \cdots = Y_n Y_{n-1} \cdots Y_1 X_0$$
$$\Rightarrow \quad \ln(X_n) = \sum_{i=1}^{n} \ln(Y_i) + \ln(X_0)$$
- Since $\ln(Y_i)$ are iid, with suitable normalization, the interpolated process $\ln(X(t))$ would be Brownian motion and $X(t)$ would be geometric Brownian motion

Gaussian Processes

- A continuous-time continuous-state process $\{X(t),\ t \geq 0\}$ is said to be a Gaussian process if for all n and all $t_1, t_2, \cdots, t_n$, we have that $X(t_1), \cdots, X(t_n)$ are jointly Gaussian.
- The Brownian motion is an example of a Gaussian Process
- The Brwonian motion is a Gaussian process with

$$E[X(t)] = 0, \quad \mathsf{Var}(X(t)) = \sigma^2 t, \quad \mathsf{Cov}(X(s), X(t)) = \sigma^2 \min(s,t)$$

- Recall that the multivariate Gaussian density is specified by the marginal means, variances and the covariances of the random variables
- Hence, a general Gaussian process is specified by the mean function and the variance and covariance functions

- Consider the statistics of the Brownian motion process for $0 < t < 1$ under the condition that $X(1) = 0$
- Consider standard Brownian motion. $(\sigma^2 = 1)$

$$E[X(t)|X(1) = 0] = \frac{t}{1}\, 0 = 0$$

Recall that, for $s < t$, conditional density of $X(s)$ conditioned on $X(t) = b$ is gaussian with mean bs/t and variance $s(t-s)/t$

Now, for $s < t < 1$, since $E[X(s)|X(1) = 0] = 0,\ s < 1$,

$$\begin{aligned}
\text{Cov}(X(s), X(t)|X(1) = 0) &\triangleq E[X(s)X(t) \mid X(1) = 0] \\
&= E[\ E[X(s)X(t) \mid X(t), X(1) = 0] \mid X(1) = 0] \\
&= E[\ X(t)E[X(s) \mid X(t)] \mid X(1) = 0] \\
&= E[\ X(t)\frac{s}{t}X(t) \mid X(1) = 0] \\
&= \frac{s}{t}\ E[X^2(t) \mid X(1) = 0] \\
&= \frac{s}{t}\ t(1-t) \\
&= s(1-t)
\end{aligned}$$

Thus, for $0 < t < 1$, conditioned on $X(1) = 0$, this process has mean 0 and covariance function $s(1-t),\ s < t$

- Consider a process $\{Z(t),\ 0 \leq t \leq 1\}$.
- It is called Brownian Bridge process if it is a Gaussian process with mean zero and covariance function $\mathrm{Cov}(Z(s), Z(t)) = s(1-t)$ when $s \leq t$.
- Let $X(t)$ be a standard Brownian motion process.
- Then, $Z(t) = X(t) - tX(1)$, $0 \leq t \leq 1$ is a Brownian Bridge
- Easy to see it is a Gaussian process with mean zero. For $s < t$

$$
\begin{aligned}
\mathrm{Cov}(Z(s), Z(t)) &= \mathrm{Cov}(X(s) - sX(1), X(t) - tX(1)) \\
&= \mathrm{Cov}(X(s), X(t)) - t\mathrm{Cov}(X(s), X(1)) - \\
&\quad s\mathrm{Cov}(X(1), X(t)) + st\mathrm{Cov}(X(1), X(1)) \\
&= s - st - st + st = s(1-t)
\end{aligned}
$$

White Noise

- Consider a process $\{V(t),\ t \geq 0\}$ with

$$E[V(t)] = 0; \quad \mathsf{Var}(V(t)) = \sigma^2 \quad \mathsf{Cov}(V(t), V(s)) = 0,\ s \neq t$$

- This is a kind of generalization of sequence of iid random variables to continuous time
- It is an example of what is called White Noise.

- Assume $V(t)$ is Gaussian. Let

$$X(t) = \int_0^t V(\tau)\, d\tau$$

- Then we get $E[X(t)] = 0$ and

$$E[X^2(t)] = \int_0^t \int_0^t E[V(t_1)V(t_2)]\, dt_1\, dt_2 = \int_0^t \sigma^2\, dt_1 = \sigma^2 t$$

$$E[X(t_1)(X(t_2) - X(t_1)] = \int_0^{t_1} \int_{t_1}^{t_2} E[V(t)V(t')]\, dt\, dt' = 0$$

- We see that $X(t)$ is a process with mean zero, variance proportional to t and having uncorrelated increments.
- One can show that it would be a Brownian motion
- The actual concept involved is rather deep

- We have considered three random processes
- Markov Chain
  – Example of Discrete-time discrete-state process
- Poisson Process
  – Example of continuous-time discrete-state process
- Brownian Motion
  – Example of continuous-time continuous-state process
- We need an example of discrete-time continuous-state process!
- Any sequence of continuous random variables would be a discrete-time continuous-state process

- Let $\{X_n,\ n = 0, 1, \cdots\}$ be a discrete-time continuous-state process.
- It is called a martingale if $E|X_n| < \infty,\ \forall n$ and

$$E[X_{n+1} \mid X_n, \cdots, X_0] = X_n,\ \forall n$$

- Suppose Z_i are iid with $Pr[Z_i = +1] = Pr[Z_i = -1] = 0.5$. Let

$$X_n = \sum_{i=1}^{n} Z_i \quad \Rightarrow \quad X_{n+1} = X_n + Z_{n+1}$$

- Since $EZ_i = 0,\ \forall i$,

$$E[X_{n+1} \mid X_n, \cdots, X_0] = E[X_n + Z_{n+1} | X_n] = X_n + E[Z_{n+1} | X_n] = X_n$$

- Hence, X_n is a martingale.
- When X_n is a martingale, we have

$$E[X_{n+1}] = E[X_n],\ \forall n$$

- $\{X_n,\ n=0,1,\cdots\}$ and $E|X_n|<\infty,\ \forall n$
- It is called a martingale if

$$E[X_{n+1}\mid X_n,\cdots,X_0]=X_n,\ \forall n$$

- It is called a supermartingale if

$$E[X_{n+1}\mid X_n,\cdots,X_0]\le X_n,\ \forall n$$

- It is called a submartingale if

$$E[X_{n+1}\mid X_n,\cdots,X_0]\ge X_n,\ \forall n$$

Please note that these are 'simplified' definitions In the above, the conditioning random variables can be another sequence Y_i if $Y_1,\cdots,Y_n$ determine $X_1,\cdots,X_n$

- Martingales are useful because of the martingale convergence theorem.

  martingale convergence theorem: If X_n is a martingale with $\sup_n E|X_n| < \infty$ then X_n converges almost surely to a rv X which will have finite expectation. A positive supermartingale also converges almost surely

- Consider the 2-armed bandit problem in problem sheet 3.7
- The algorithm is

$$\begin{aligned} p(k+1) &= p(k) + \lambda(1-p(k)) \;\; \text{if arm 1 chosen, } b(k)=1 \\ &= p(k) - \lambda p(k) \;\; \text{if arm 2 is chosen and } b(k)=1 \\ &= p(k) \;\; \text{if } b(k)=0 \end{aligned}$$

- We get

$$\begin{aligned} & E[p(k+1)-p(k)|p(k)] \\ & \quad = \lambda(1-p(k)) \; Pr[b(k)=1, \text{arm 1} \mid p(k)] \\ & \qquad -\lambda p(k) \; Pr[b(k)=1, \text{arm 2} \mid p(k)] \\ & \quad = \lambda(1-p(k)) \; Pr[b(k)=1 \mid \text{arm 1}, p(k)] \; Pr[\text{arm 1} \mid p(k)] \\ & \qquad -\lambda p(k) \; Pr[b(k)=1 \mid \text{arm 2}, p(k)] \; Pr[\text{arm 2} \mid p(k)] \end{aligned}$$

- This gives us

$$\begin{aligned} E[p(k+1) - p(k)|p(k)] &= \lambda(1 - p(k))\ d_1\ p(k) \\ &\quad -\lambda p(k)\ d_2\ (1 - p(k) \\ &= \lambda\, p(k)(1 - p(k))\ (d_1 - d_2) \\ &\geq 0, \quad \text{if} \quad d_1 > d_2 \end{aligned}$$

$$\Rightarrow \quad E[p(k{+}1) \mid p(k)] \geq p(k) \quad \Rightarrow E[p(k{+}1)] \geq E[p(k)],\ \forall k$$

- This also shows $p(k)$ is a submartingale.
- Here, $p(k)$ is bounded and $1 - p(k)$ is a supermartingale.
- So, we can conclude, the algorithm converges almost surely

- We have mentioned martingales as an example of discrete-time continuous processes
- A stochastic iterative algorithm essentially generates a discrete-time continuous-state processes.
- Martingales are very useful in analyzing convergence of many stochastic algorithms
- While we mentioned only discrete-time martingales, one can similarly have continuous-time martingales