## The normal distribution

A random variable $Z$ has the normal distribution with mean $\mu$ and variance $\sigma^2$ denoted by $Z \sim \mathrm{N}(\mu, \sigma^2)$, if its density is

$$\phi_{\mu,\sigma^2}(z) = \frac{\exp\{-(z-\mu)^2/(2\sigma^2)\}}{\sqrt{2\pi}\sigma},$$

and then its distribution function is

$$\Phi_{\mu,\sigma^2}(z) = \int_{-\infty}^{z} \phi_{\mu,\sigma^2}(t)dt.$$

[We'll drop the subscripts when $\mu = 0$ and $\sigma^2 = 1$]
If $Z \sim \mathrm{N}(0,1)$, then $(\sigma Z + \mu) \sim \mathrm{N}(\mu, \sigma^2)$.

# Brownian motion

The normal distribution arises as the limit of random walks.

- If $X_1, X_2, \ldots$ are i.i.d. with mean 0 and variance 1, then for $S_n = \sum_{i=1}^n X_i$,

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sqrt{n}} \leq z\right) = \Phi(z).$$

Moreover, we can also sum a different number of terms, but keep the scaling the same: If $t \geq 0$ then

$$\lim_{n\to\infty} P\left(\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \leq z\right) = \Phi_{0,t}(z).$$

# Brownian motion

**Definition:**

A continuous time stochastic process $\{B_t : t \geq 0\}$ is a standard Brownian motion if

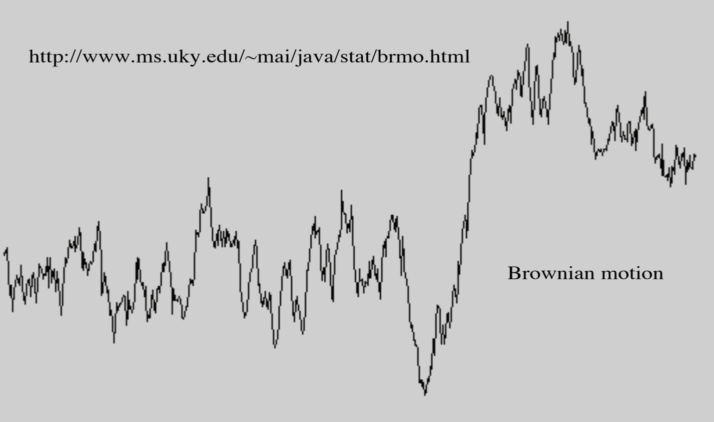
- it has continuous sample paths,
- it has independent increments on disjoint intervals: for $k \geq 2$ and $0 \leq s_1 < t_1 \leq s_2 < \cdots < t_k$,

$$B_{t_1} - B_{s_1}, \ldots, B_{t_k} - B_{s_k}$$

  are independent variables.
- For each $t \geq 0$, $B_t \sim \mathrm{N}(0, t)$.

# Brownian Motion

## A sample path



# Properties of Brownian motion

- $B_{t+s} - B_t \sim \mathrm{N}(0, s)$.
- Furthermore if, for fixed $h$, we define $B^*_t = B_{t+h} - B_h$, then $B^*$ is a standard Brownian motion.
- Brownian motion with parameter $\sigma^2$ is defined to have the same distribution as $(\sigma B_t)_{t\geq 0}$.

# Joint distributions of Brownian motion

Let $0 = t_0 < t_1 < \cdots < t_k$. What is the joint distribution of $(B_{t_1}, \ldots, B_{t_k})$?

- Let $Z_i = B_{t_i} - B_{t_{i-1}}$, $i = 1, \ldots, k$. The $Z_i$ are independent normal random variables.
- The $B_{t_i}$ are a linear function of the $Z_i$:

$$(B_{t_1}, \ldots, B_{t_k}) = \left( Z_1, \sum_{i=1}^{2} Z_i, \ldots, \sum_{i=1}^{k} Z_i \right).$$

The joint distributions of Brownian motion observed at a collection of times are linear functions of independent normal variables. What are these distributions?

# Multivariate normal distribution

To define the Multivariate normal distribution we need some facts from linear algebra.

**Definition**
We say the matrix $\Sigma$ is positive definite if $\Sigma^T = \Sigma$ and for any $x \neq 0$, $x^T \Sigma x > 0$.

**Properties of positive definite matrix $\Sigma$**

- There is a lower triangular matrix $R$ with $\Sigma = RR^T$.
- There is a unique symmetric square root denoted $\Sigma^{1/2}$.
- $\det(\Sigma) > 0$. (In particular $\Sigma$ is invertible.)

# Multivariate normal distribution

Let $Z = (Z_1, \ldots, Z_k)$ be a vector of i.i.d. standard normal variables.

**Definition**

We say $X = (X_1, \ldots, X_k)$ has the multivariate normal distribution with parameters $\mu$, a $k$-vector called the mean, and $\Sigma$, a $k \times k$ positive definite matrix called the variance or covariance matrix, if

$$X \overset{d}{=} \Sigma^{1/2} Z + \mu.$$

# Multivariate normal distribution

**Properties**

- $Cov(X_i, X_j) = \Sigma_{i,j}$. [Direct calculation.]
- If $R$ is such that $\Sigma = RR^T$ (and there is such a lower triangular matrix $R$), then

$$X \stackrel{d}{=} RZ + \mu.$$

  [Check densities.]
- If $A$ is an invertible matrix, then $AX$ is multivariate normal with mean $A\mu$ and covariance matrix $A\Sigma A^T$. [Use second item, checking the covariance matrix is positive definite.]

# Multivariate normal distribution

- The third point says that if $X = AZ + \mu$ for a an invertible matrix $A$ then $X$ is multivariate normal with covariance matrix $AA^T$.
- Alternatively the first item says that once it's established that $X$ is multivariate normal (e.g., by recognizing it as a linear function of a multivariate normal vector), then the covariance matrix has $(i, j)$th entry $Cov(X_i, X_j)$.

# Multivariate normal distribution

The density of $X$ is

$$f(x) = \frac{1}{(2\pi)^{k/2}\sqrt{\det(\Sigma)}} \exp\left\{-(x-\mu)^T\Sigma^{-1}(x-\mu)/2\right\}.$$

This expression is difficult to compute with in practice so it's best to use a convenient representation as a linear function of independent normal variables.

# Multivariate normal distribution

**Example: bivariate normal**

$(X_1, X_2)$ are bivariate normal with correlation $\rho$ and means $\mu_1, \mu_2$ and variances $\sigma_1^2, \sigma_2^2$.

- What are the multivariate normal parameters $\mu$ and $\Sigma$ in terms of the parameters above?
- Find a lower triangular $R$ such that $RR^T = \Sigma$ and $X = RZ + \mu$.
- Write down the joint density of $(X_1, X_2)$.

# Joint distribution of Brownian motion

- We saw that Brownian motion observed at a collection of times are linear functions of independent normal distributions and so are distributed as multivariate normal.
- The means are zero and the so the distribution is entirely determined by the pairwise covariances.

We can compute the covariance of Brownian motion observed at times $s < t$.

$$\begin{aligned}
Cov(B_t, B_s) &= E[B_t B_s] && (E[B_t]{=}0) \\
&= E[(B_t - B_s)B_s] + E[B_s^2] \\
&= E[B_t - B_s]E[B_s] + Var(B_s) && \text{(ind. incs.)} \\
&= s.
\end{aligned}$$

# Joint distribution of Brownian motion

If $0 < t_1 < \cdots < t_k$ then $(B_{t_1}, \ldots, B_{t_k})$ is multivariate normal with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & \cdots & t_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & \cdots & t_k \end{pmatrix}.$$

# Finance example

Assume that the logarithm of the (standardized) price of stock $t$ hours into the trading day is given by $\sigma B_t$ for some $\sigma > 0$ and where $B_t$ is a Brownian motion.

- If the stock is worth $e^{4\sigma}$ dollars halfway through the 8 hour trading day, what is the chance the stock will be worth more than its initial price at the end of the day?
- If at the end of the day the stock is worth $e^{4\sigma}$ dollars, what is the chance the stock's price at the middle of the day was greater than its starting price?

# Properties of Brownian motion

Brownian motion arises as the limit of random walk, inherits definition/properties from there.

- This result is called Donsker's Theorem or the invariance principle.
- Is a good way to approximately simulate Brownian motion.

We can use the ideas around limits of discrete processes and generators to understand this result a bit more.

# Einstein derivation of Brownian motion

- We'll derive a PDE for $p_t(x)$, the density of the limit of simple random walk (properly scaled).
- A solution to this PDE is

$$p_t(x) = \frac{\exp[-x^2/(2t)]}{\sqrt{2\pi t}},$$

which is the density of $\mathrm{N}(0, t)$.

# Einstein derivation of Brownian motion

$N$ is a (large) integer.

- Let $X_1^{(N)}, X_2^{(N)}, \ldots$ be i.i.d. with

$$P\left(X_i^{(N)} = \frac{\pm 1}{\sqrt{N}}\right) = 1/2.$$

- Simple random walk with jumps at times $1/N, 2/N, \ldots$:

$$S_t^{(N)} = \sum_{i=1}^{\lfloor Nt \rfloor} X_i^{(N)}.$$

- $Q_{n/N}(k/\sqrt{N}) = P\left(S_{n/N}^{(N)} = k/\sqrt{N}\right)$.

# Einstein derivation of Brownian motion

- $Q_{n/N}(k/\sqrt{N}) = P\left(S^{(N)}_{n/N} = k/\sqrt{N}\right)$.

If we let $n$ and $k$ grow with $N$ such that as $N \to \infty$,

$$n/N \to t > 0 \quad \text{and} \quad k/\sqrt{N} \to x,$$

then, by the CLT,

$$\sqrt{N} Q_{n/N}(k/\sqrt{N}) \to p_t(x).$$

# Einstein derivation of Brownian motion

By the law of total probability:

$$Q_{\frac{n}{N}}\left(\frac{k}{\sqrt{N}}\right) = \frac{1}{2}Q_{\frac{n-1}{N}}\left(\frac{k+1}{\sqrt{N}}\right) + \frac{1}{2}Q_{\frac{n-1}{N}}\left(\frac{k-1}{\sqrt{N}}\right),$$

and so

$$\begin{aligned} N&\left[Q_{\frac{n}{N}}\left(\frac{k}{\sqrt{N}}\right) - Q_{\frac{n-1}{N}}\left(\frac{k}{\sqrt{N}}\right)\right] \\ &= \frac{N}{2}\left[Q_{\frac{n-1}{N}}\left(\frac{k+1}{\sqrt{N}}\right) - Q_{\frac{n-1}{N}}\left(\frac{k}{\sqrt{N}}\right)\right] \\ &\quad - \frac{N}{2}\left[Q_{\frac{n-1}{N}}\left(\frac{k}{\sqrt{N}}\right) - Q_{\frac{n-1}{N}}\left(\frac{k-1}{\sqrt{N}}\right)\right]. \end{aligned}$$

As $N \to \infty$, remembering that $k/\sqrt{N} \to x$ and $n/N \to t$,

- LHS$\to \frac{\partial}{\partial t} p_t(x)$,
- RHS$\to \frac{1}{2}\frac{\partial^2}{\partial x^2} p_t(x)$.

# Einstein derivation of Brownian motion

So the limiting stochastic process should have density $p_t(x)$ at time $t$ satisfying the PDE

$$\frac{\partial}{\partial t} p_t(x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} p_t(x).$$

This PDE is called the heat equation and under appropriate boundary conditions the unique solution is

$$p_t(x) = \frac{\exp[-x^2/(2t)]}{\sqrt{2\pi t}}.$$

We can think of the heat equation for Brownian motion as the continuous state space analog of the forward equation for CTMCs:

$$\frac{\partial}{\partial t} p_t = \mathcal{A}(p_t),$$

where $\mathcal{A}$ is the linear operator on twice differentiable functions with

$$\mathcal{A}f(x) = \frac{1}{2} f''(x).$$

# Hitting times of Brownian motion

- Define the hitting time of level $x$ by $T_x = \inf\{t : B_t = x\}$.
- Brownian motion is continuous so if $0 < x < y$, then $T_x < T_y$.
- Since simple symmetric random walk is recurrent, $T_x$ is finite.

# Hitting times of Brownian motion

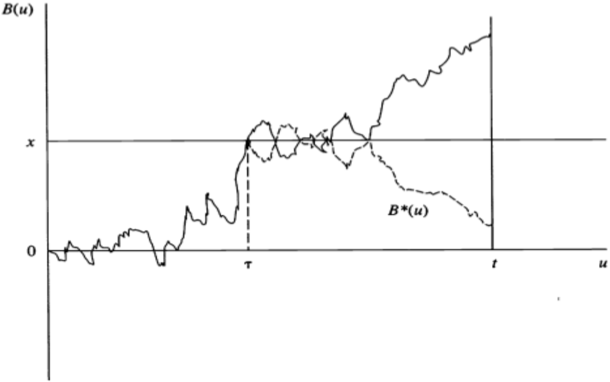
We derive the distribution of $T_x$ for $x > 0$ by using the relation:

$$\begin{aligned} P(B_t \geq x) &= P(B_t \geq x, T_x \leq t) \\ &= P(B_t \geq x | T_x \leq t) P(T_x \leq t). \end{aligned}$$

We only need to determine $P(B_t \geq x | T_x \leq t)$ since we know the distribution of $B_t$ and hence the LHS.

# Reflection principle

If $\tau < t$, then $P(B_t - x > 0 | T_x = \tau) = 1/2$.



# Hitting times of Brownian motion

Combining the last two slides we have the distribution function

$$\begin{aligned} P(T_x \le t) &= 2P(B_t > x) \\ &= \sqrt{\frac{2}{\pi t}} \int_x^\infty \exp[-u^2/(2t)]du \\ &= \sqrt{\frac{2}{\pi}} \int_{x/\sqrt{t}}^\infty \exp[-u^2/2]du. \end{aligned}$$

So $T_x$ has density

$$f_{T_x}(t) = \frac{xt^{-3/2}}{\sqrt{2\pi}} \exp[-x^2/(2t)].$$

# The Maximum of Brownian motion on an interval

Let $M_t = \max_{0 \le s \le t} B_s$. The distribution of $M_t$ is now easily derived. For $x > 0$,

$$\begin{aligned} P(M_t \le x) &= P(T_x > t) \\ &= 1 - 2P(B_t > x) \\ &= P(-x \le B_t \le x) \\ &= P(|B_t| \le x). \end{aligned}$$

So for each fixed $t$, $M_t \stackrel{d}{=} |B_t|$ (but not as processes!), and the maximum of Brownian motion is distributed as the absolute value of a normal distribution.

# Gambler's ruin via the invariance principle

Let $x < 0 < y$. What is $P(T_x < T_y)$?

- Gambler's ruin from DTMC slides says that the chance simple symmetric random walk hits $-L$ before hitting $M$ is $M/(L+M)$.
- Using the approximation from before, we set $-L = \lfloor \sqrt{N} x \rfloor$ and $M = \lfloor \sqrt{N} y \rfloor$ and take the limit as $N \to \infty$.

$$P(T_x < T_y) = \frac{y}{y - x}.$$