Stochastic Calculus: NYU, Fall 2003

1 Preliminaries

1.1 Multivariate Normals

Definition 1 *A Gaussian vector (multivariate normal random variable),* $X = \begin{pmatrix} X_1 \\ ... \\ X_p \end{pmatrix}$*, is a random variable in* R^p *with the density function* $f_X(x) = \frac{1}{(2\pi)^{p/2}|\det\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$*, where* $\mu \in R^p$ *is the mean and* $\Sigma \in R^{p\times p}$ *is a symmetric positive definite matrix and the covariance matrix.*

Note that the fact that Σ is positive definite (and not positive semidefinite) implies that no linear combination of some of the components is perfectly correlated with the other components. Some other properties include:

- Each coordinate is Gaussian.
- Each subset of the coordinates with also a Gaussian vector.
- If C is a non-singular $p\times p$ matrix and $Y = m+CX$, then Y is distributed multivariate normal with mean $m + C\mu$ and covariance $C\Sigma C^T$.

2 Limit Theorems

Theorem 2 *(Weak Law of Large Numbers). Let* ξ_j *be a sequence of independent, identically distributed random variables. Let* $\eta = E(\xi_j) < \infty$*. Let* $S_n = \sum_{j=1}^n \xi_j$*. Then,* $P(\left|\frac{S_n}{n} - \eta\right| \le \varepsilon) \longrightarrow 1$ *as* $n \longrightarrow \infty$ *for all* $\varepsilon > 0$*.*

Proof. For simplicity, we assume that $\sigma^2 = E(\xi_j^2) < \infty$. Without loss of generality, assume that $\eta = 0$ (if not, we replace ξ_j by $\xi_j - \eta$). By Chebyshev's Inequality, $P(|X| \ge \varepsilon) \le \frac{1}{\varepsilon^p}E(|X|^p)$ for all $p > 0$, since $E(|X|^p) = \int_R |X|^p \mu(dx) \ge \int_{|X|\ge\varepsilon} |X|^p \mu(dx) \ge \varepsilon^p \int_{|X|\ge\varepsilon} \mu(dx) = \varepsilon^p P(|X| \ge \varepsilon)$. Applying Chebyshev's Inequality, we find:

$$P(\left|\frac{S_n}{n}\right| \ge \varepsilon) \le \frac{1}{\varepsilon^2}E(\left|\frac{S_n}{n}\right|^2) = \frac{1}{n^2}\frac{1}{\varepsilon^2}E(|S_n|^2)$$

Using the fact that the ξ_j are independent and identically distributed, we find:

$$\begin{aligned} E(|S_n|^2) &= E((\sum_{j=1}^n \xi_j)^2) \\ &= E(\sum_{j=1}^n {\xi_j}^2) \\ &= n\sigma^2 \end{aligned}$$

Thus, $P(\left|\frac{S_n}{n}\right| \geq \varepsilon) \leq \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$, which goes to 0 for fixed ε as $n \longrightarrow \infty$. Notice that $P(\left|\frac{S_n}{n}\right| \leq \varepsilon) \longrightarrow 1$ if and only if $P(\left|\frac{S_n}{n}\right| > \varepsilon) \longrightarrow 0$. Thus, $P(\left|\frac{S_n}{n} - \eta\right| \leq \varepsilon) \longrightarrow 1$ as $n \longrightarrow \infty$ for all $\varepsilon > 0$. ■

Theorem 3 *(Central Limit Theorem) Let ξ_j be a sequence of independent, identically distributed random variables. Let $\eta = E(\xi_j) < \infty$. Let $S_n = \sum_{j=1}^{n} \xi_j$. Let $E(\xi_j^2) = \sigma^2 < \infty$. As $n \longrightarrow \infty$, $\frac{S_n - n\eta}{\sqrt{n\sigma^2}}$ converges is probability to a normal random variable with mean 0 and variance 1, that is, a random variable with probability density $\rho(x) = e^{-x^2/2}/\sqrt{2\pi}$.*

Proof. We use the characteristic function of x: $f(z) = E(e^{izx}) = \int_R e^{izx}\rho_x(x)dx$. Note that the characteristic function is unique to a density, and that the characteristic function of a standard normal random variable is $f_G(z) = e^{-z^2/2}$. Without loss of generality, we assume that $\eta = 0$ (if not, we replace ξ_j by $\xi_j - \eta$). We find the characteristic function of $S_n/\sqrt{n\sigma^2}$ using a Taylor expansion and the definition of e:

$$
\begin{aligned}
g_n(z) &= E(\prod_{j=1}^{n} e^{\frac{i}{\sqrt{n\sigma^2}}\xi_j z}) \\
&= \prod_{j=1}^{n} E(e^{\frac{i}{\sqrt{n\sigma^2}}\xi_j z}) \\
&= \prod_{j=1}^{n} E(1 + (\frac{iz}{\sqrt{n\sigma^2}}\xi_j) - \frac{z^2}{2(\sqrt{n\sigma^2})^2}\xi_j^2 + ...) \\
&= \prod_{j=1}^{n}(1 + 0 + -\frac{z^2}{2n\sigma^2}\sigma^2 + ...) \\
&\approx \prod_{j=1}^{n}(1 - \frac{z^2}{2n}) \\
&= (1 - \frac{z^2}{2n})^n \\
&\longrightarrow e^{-z^2/2}
\end{aligned}
$$

Since the characteristic function converges to the characteristic function of a normal random variable, $\frac{S_n - n\eta}{\sqrt{n\sigma^2}}$ converges to a random variable with a standard normal distribution. ■

Lemma 4 *(Borel-Cantelli Lemma) Let s be a sequence of events. Let $B_j = \{w : (w_{j+1}, ..., w_{j+k}) = s\}$, where each w is an infinite vector (string) of events. Then, $P(B_j$ infinitely often$) = 0$ if $\sum_{j=1}^{\infty} P(B_j) < \infty$ and $P(B_j$ infinitely often$) = 1$ if the B_j are independent and $\sum_{j=1}^{\infty} P(B_j) = \infty$.*

Proof. B_j occurs infinitely often if $P(\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty} B_j) > 0$. Note that $P(\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty} B_j) \leq P(\bigcup_{j=k}^{\infty} B_j) \leq \sum_{j=k}^{\infty} P(B_j)$. In the first case, $\sum_{j=1}^{\infty} P(B_j)$ converges, so that $\sum_{j=k}^{\infty} P(B_j) \longrightarrow 0$ as $k \longrightarrow \infty$. Since this is an upper bound on $P(B_j$ infinitely often), $P(B_j$ infinitely often$) = 0$. In the second case, ■

Theorem 5 *(Strong Law of Large Numbers) Let* $\{X_j\}_{j=1}^{\infty}$ *be a sequence of independent, identically distributed random variables. Let* $\eta = E(X)$. *Let* $S_n = \sum_{j=1}^{n} X_j$. *Then,* $\frac{S_n}{n} \longrightarrow \eta$ *almost surely if and only if* $E(|X_j|) < \infty$.

Proof. For simplicity, we assume that $E(X_j^4) < \infty$, which also implies that $E(X_j^2) < \infty$. Using the Chebyshev inequality with $p = 4$, we find that:

$$P(\left|\frac{S_n}{n}\right| > \varepsilon) \leq \frac{1}{n^4\varepsilon^4}E(S_n^4)$$

$$\begin{aligned} E(S_n^4) &= E(\sum\sum\sum\sum \xi_{j_1}\xi_{j_2}\xi_{j_3}\xi_{j_4}) \\ &= \sum_{j=1}^{n} E(\xi_j^4) + 3\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} E(\xi_i^2)E(\xi_j^2) \\ &= nE(\xi_j^4) + 3n(n-1)(E(\xi_i^2))^2 \end{aligned}$$

Thus, $E(S_n^4)$ grows as n^2, since $E(\xi_j^4)$ and $E(\xi_i^2)$ are both fixed and finite. Thus, $P(\left|\frac{S_n}{n}\right| > \varepsilon)$ is bounded above by a function of order $\frac{1}{n^2}$ for any fixed ε. Choose $B_n = \{w : w$ yields a sum of S_n with $\left|\frac{S_n}{n}\right| > \varepsilon)$. Then, $\sum_{j=1}^{\infty} P(B_j)$ is the sum of numbers bounded above by $\frac{c}{n^2}$ for some c. Since $\sum_{n=1}^{\infty} \frac{c}{n^2}$ converges, $\sum_{j=1}^{\infty} P(B_j) < \infty$. Hence, $P(\left|\frac{S_n}{n}\right| > \varepsilon$ infinitely often$) = 0$ and $\frac{S_n}{n} \longrightarrow \eta$ almost surely. ■

2.1 Statistics of Extrema

Theorem 6 *Let* $\{\xi_j\}_{j\in N}$ *be a sequence of independent, identically distributed random variables. Let* $M_n = \max\{\xi_1, ..., \xi_n\}$. *If there exist* a_n, b_n *such that* $P(a_n(M_n - b_n) \leq x) \longrightarrow G(x)$ *as* $n \longrightarrow \infty$, *then* $G(x)$ *is of one of three forms: (1)* $G(x) = e^{-e^{-x}}$, *(2)* $G(x) = e^{-x^{-\alpha}}, x > 0, \alpha > 0$, *or (3)* $G(x) = e^{-|x|^{\alpha}}, x \leq 0, \alpha > 0$.

3 Markov Chains

Definition 7 *Given a sequence* $\{X_n\}_{n\in N}$, *it has the Markov property if* $P(X_n = i|X_{n-1} = j_{n-1}) = P(X_n = i|X_{n-1} = j_{n-1}, X_{n-2} = j_{n-2}, ...)$. *If a sequence has the Markov property, we call it a Markov chain. We may then specify its evolution through the transition probabilities* $p_{ij}^{(n)} = P(X_n = i|X_{n-1} = j)$.

Definition 8 *A process is stationary if* $p_{ij}^{(n)} = P(X_n = i|X_{n-1} = j) = p(i|j)$ *does not depend on* n*. Then, we may simply write* $p_{ij} = p_{ij}^{(n)}$*. This also defines a matrix of transition probabilities,* P*, with each row sum equal to 1.*

Note that defining an initial distribution, for X_0, and the transition probabilities completely defines a Markov process.

Some properties of Markov chains with a state space, S, initial distribution, μ, and transition probabilities $p(i|j)$ include:

- $\sum_{i \in S} \mu(i) = 1$ (there is an initial condition in the space)
- For all $j \in S$, $\sum_{i \in S} p(i|j) = 1$ (there is always a transition, even if it is to the same state)
- $0 \leq \mu(i) \leq 1$ and $0 \leq p(i|j) \leq 1$ (these are probabilities)

Definition 9 *We say that* i *leads to* j*,* $i \longrightarrow j$*, if there exists* s *such that the* ji *entry of* P^s *is positive for some* s*. That is, there is some chain of finite length and non-zero probability from* i *to* j*. If* $i \longrightarrow j$ *and* $j \longrightarrow i$*, then we say that* i *and* j *communicate, and* $i \leftrightarrow j$*.*

Theorem 10 *If there exists a fixed* s *such that for all* (i, j) $P_{ij}^s > 0$*, then (1) there exists a unique* π *such that* $\pi = P\pi$ *and (2) for all* μ_0 *then* $\mu_n = P^n \mu_0$ *converges to* π *as* $n \longrightarrow \infty$ *exponentially fast.*

Proof. Insert proof here. ■

Definition 11 *A chain is ergodic if all pairs of states communicate. That is, there are no disconnected chains or groups of states to which return with probability 0.*

To find the probabilities of certain events, such as entering a state or getting out of a certain set of states, we may create a modified chain, with a black hole state (one from which one cannot exit) that is entered only when that event occurs. Then, the probability of that event already having occurredin the original chain equals the probability of being in the black hole state in the modified chain.

4 Continuous Time Stochastic Processes

4.1 Brownian Motion (the Weiner Process)

Let $\{\xi_n\}_{n \in N}$ be independent and identically distributed with $P(\xi_n = +1) = P(\xi_n = -1) = \frac{1}{2}$. Then, $S_n = \sum_{i=1}^{n} \xi_i$ is a random walk, and we may write $S_n = S_{n-1} + \xi_n$. For any integer N, set $\Delta t = 1/N$ for $t \in [0, 1]$. Set $x_{k/N} = \frac{S_k}{\sqrt{N}}$. We may define a piecewise continuous function by $x_N(t) = x_{\lfloor tN \rfloor / N}$. Under appropriate assumptions (Donsker), $x_N(t)$ converges in distribution to a random process, W_t. We call this the Weiner Process, or Brownian motion. Some properties of this process include:

- $W_1 \tilde{} Normal(0,1)$

- $W_t \tilde{} Normal(0,t)$, since $X_t^N = \lim \frac{S_{\lfloor Nt \rfloor}}{\sqrt{N}} = \lim \frac{S_{\lfloor Nt \rfloor}}{\sqrt{\lfloor Nt \rfloor}} \frac{\sqrt{\lfloor Nt \rfloor}}{\sqrt{N}}$, where the first term converges in distribution to a standard normal random variable and the second converges to $\sqrt{t}$.

- Since $\{S_n\}_{n \in N}$ is Markov, the Weiner process is a continuous-time Markov process. That is, $P(W_t \leq x | \{W_{s'}\}_{s' \leq s}) = P(W_t \leq x | W_s)$.

Using the Markov property, we find that $P(W_t \leq x | W_s = y) = P(W_{t-s} \leq x - y)$, and $W_t - W_s$ has the same distribution as W_{t-s}, that is, $Normal(0, t-s)$. Extending this, we find that, for any partition $\{t_0 = 0, t_1, ..., t_{n-1}, t_n = 1\}$ of $[0,1]$, the joint probability density function of $W_{t_n}, ..., W_{t_0} = 0$ is $\rho_{t_n - t_{n-1}}(x_n | x_{n-1}) ... \rho_{t_2 - t_1}(x_2 | x_1) \rho_{t_1 - 0}(x_1 | 0) = \rho_{t_n - t_{n-1}}(x_n - x_{n-1} | 0) ... \rho_{t_2 - t_1}(x_2 - x_1 | 0) \rho_{t_1 - 0}(x_1 | 0)$.

The covariance of any two points in time, $t > s$, is given by:

$$\begin{aligned}
E(W_t W_s) &= \int_{R^2} xy \rho_{t-s}(x|y) \rho_s(y|0) dx dy \\
&= \int_{R^2} xy \frac{e^{-(x-y)^2/2(t-s)}}{\sqrt{2\pi(t-s)}} \frac{e^{-y^2/2s}}{\sqrt{2\pi s}} dx dy \\
&= \int_{R^2} (x-y) y \frac{e^{-(x-y)^2/2(t-s)}}{\sqrt{2\pi(t-s)}} \frac{e^{-y^2/2s}}{\sqrt{2\pi s}} dx dy + \int_{R^2} y^2 \frac{e^{-(x-y)^2/2(t-s)}}{\sqrt{2\pi(t-s)}} \frac{e^{-y^2/2s}}{\sqrt{2\pi s}} dx dy \\
&= 0 + E(W_s^2) = \min(s,t)
\end{aligned}$$

Thus, $E(W_s W_t) = \min(s,t)$. In addition,

$$\begin{aligned}
E((W_t - W_s)^2) &= E(W_t^2) + E(W_s^2) - 2E(W_s W_t) \\
&= t + s - 2\min(t,s) \\
&= |t - s|
\end{aligned}$$

This implies that the Weiner process is almost surely continuous, since $\lim_{t \to s} E((W_t - W_s)^2) = 0$. However, it is almost surely not differentiable anywhere, since $E((W_t - W_s)^2)$ grows as $(t-s)^2$.

The Weiner Process is self-similar: Given $\lambda > 0$, $\lambda^{-1/2} W_{\lambda t}$ is equal in distribution to W_t, since $E(\lambda^{-1/2} W_{\lambda t} \lambda^{-1/2} W_{\lambda s}) = \lambda^{-1} E(W_{\lambda t} W_{\lambda s}) = \lambda^{-1} \min(\lambda t, \lambda s) = \min(s,t)$. This allows us to study the properties of the Weiner process on $[0,1]$ and apply them to any interval.

Another construction of the Weiner process: Let $\{f_k(t)\}$ be an orthonormal basis (so that $\int_0^1 f_k(t) f_{k'}(t) dt = \delta_{k,k'}$) in $L^2[0,1]$, so that for any $g : [0,1] \to R$, with $\int_0^1 g(t)^2 dt < \infty$, we may write $g(t) = \sum_k \alpha_k f_k(t)$, where $\alpha_k = \int_0^1 g(t) f_k(t) dt$. Note that $\int_0^1 g(t)^2 dt = \sum_k \alpha_k^2$. Let $\{\beta_k\}$ be indiependent, identically distributed standard normal random variables. Let $W_t = \sum_k \beta_k \int_0^t f_k(z) dz$. Then, since each $\int_0^t f_k(z) dz$ is fixed, each W_t is normal with mean 0. In addition,

$$\begin{aligned} E(W_t W_s) &= E(\sum_k \sum_{k'} \beta_k \beta_{k'} \int_0^t f_k(z)dz \int_0^s f_{k'}(z')dz') \\ &= \sum_k \sum_{k'} E(\beta_k \beta_{k'}) \int_0^t f_k(z)dz \int_0^s f_{k'}(z')dz' \\ &= \sum_k \int_0^t f_k(z)dz \int_0^s f_k(z')dz' \end{aligned}$$

Let $\chi_t(z)$ be the indicator function for the interval $[0,t]$. Since $\chi_t(z) \in L^2[0,1]$, we may write $\chi_t(z) = \sum_k f_k(t)(\int_0^1 \chi_t(z')f_k(z')dz') = \sum_k f_k(t)(\int_0^t \chi_t(z')f_k(z')dz')$, so that $\sum_k \int_0^t f_k(z)dz \int_0^s f_k(z')dz' = \int_0^1 \chi_t(z)\chi_s(z)dz$ In addition, $\int_0^1 \chi_t(z)\chi_s(z)dz = \int_0^{\min(t,s)} \chi_t(z)\chi_s(z)dz = \min(t,s)$. Thus, $E(W_t W_s) = \min(t,s)$.

4.2 Filtrations and Martingales

Definition 12 *A σ-field, F, on a probability space, Ω, is a collection of subsets of Ω which contains $\emptyset$, Ω, and is closed under completements and countable unions and intersections. For a stochastic process $Y = (Y_t, t \in [0,T], \omega \in \Omega)$, the σ-field $\sigma(Y)$ is the smallest σ-field containing all sets of the form $\{\omega : (Y_t, t \in [0,T]) \in C\}$, where C is a set of functions on $[0,T]$. This is the σ-field generated by Y.*

Basically, $\sigma(Y)$ contains all information about the structure of Y. In particular, if $s < t$, then $\sigma(Y_s) \subset \sigma(Y_t)$, because more information is known about the path of Y at a later time.

Definition 13 *A collection $(F_t, t \geq 0)$ of σ-fileds of Ω is called a filtration if $F_s \subset F_t$ for all $0 \leq s \leq t$. The stochastic process Y_t is adapted to the filtration $(F_t, t \geq 0)$ if $\sigma(Y_t) \subset F_t$ for all $t \geq 0$. The natural filtration generated by a stochastic process Y_t is $F_t = \sigma(Y_s, s \leq t)$.*

Definition 14 *The stochastic process X_t is called a continuous-time martingale with respect to the filtration (F_t) if $E|X_t| < \infty$, X_t is adapted to (F_t), and $E(X_t|F_s) = X_s$ for all $0 \leq s < t$.*

Brownian motion is a martingale.

4.3 Gaussian Processes

Definition 15 *X_t, for $t \in [0,1]$, is a Gaussian process if, for any partition $0 < t_1 < ... < t_n \leq 1$, the vector $\begin{pmatrix} X_{t_1} \\ ... \\ X_{t_n} \end{pmatrix}$ is a Gaussian vector.*

As with Gaussian normals, Gaussian processes are completely determined by their mean and covariance. In this case, X_t is completely determined by $E(X_t)$ and $E(X_t X_s) = K(t,s)$ for all $t, s \in [0,1]$. Given a Gaussian process, G_t, with a zero mean, we may construct a process with mean m_t at each time as $G_t + m_t$.

Definition 16 *W_t is the Weiner process (Brownian motion) if (1) W_t is a Gaussian process, (2) $E(W_t) = 0$ and $E(W_s W_t) = \min(s,t)$, and (3) W_t is almost surely continuous.*

Given a certain covariance, $K(t,s)$, we may construct a zero mean Gaussian process on $[0,1]$ using the Karhunen-Loeve Expansion. First, we find a countable set of functions, $\{\varphi_k\}_{k\in N}$, $\varphi_k : [0,1] \to R$, such that $\int_0^1 K(t,s)\varphi_k(s)ds = \lambda_k \varphi_k(t)$ for each k. (We also assume that we have the ordering $\lambda_1 > \lambda_2 > ... > 0$.) Then, we may write $K(t,s) = \sum_{k=0}^{\infty} \lambda_k \varphi_k(t)\varphi_k(s)$. Then, the Gaussian process with covariance $K(t,s)$ is $G_t = \sum_{k=0}^{\infty} \sqrt{\lambda_k}\xi_k \varphi_k(t)$, where $\{\xi_k\}_{k\in N}$ are independent, identically distributed standard Gaussian random variables. Clearly, this process is Gaussian (since each point is a linear combination of independent random variables), and its covariance is actually $K(t,s)$:

$$
\begin{aligned}
E(G_t G_s) &= E(\sum_k \sum_{k'} \sqrt{\lambda_k \lambda_{k'}} \xi_k \xi_{k'} \varphi_k(t) \varphi_{k'}(s)) \\
&= \sum_k \sum_{k'} \sqrt{\lambda_k \lambda_{k'}} \varphi_k(t)\varphi_{k'}(s) E(\xi_k \xi_{k'}) \\
&= \sum_k \lambda_k \varphi_k(t)\varphi_k(s) \\
&= K(t,s)
\end{aligned}
$$

We find the Karhunen-Loeve expansion of the Weiner process. In this case, $K(t,s) = \min(t,s)$. Thus, we solve:

$$
\begin{aligned}
\lambda\varphi(t) &= \int_0^1 \min(t,s)\varphi(s)ds \\
&= \int_0^t \min(t,s)\varphi(s)ds + \int_t^1 \min(t,s)\varphi(s)ds \\
&= \int_0^t s\varphi(s)ds + \int_t^1 t\varphi(s)ds
\end{aligned}
$$

Taking the first and second derivatives with respect to t, we find:

$$
\begin{aligned}
\lambda\varphi'(t) &= t\varphi(t) - t\varphi(t) + \int_t^1 \varphi(s)ds \\
&= \int_t^1 \varphi(s)ds \\
\lambda\varphi''(t) &= -\varphi(t)
\end{aligned}
$$

The general solution to this differential equation is $\varphi(t) = A\sin(\frac{t}{\sqrt{\lambda}} + B)$. From the original equation, we have the boundary condition that $\varphi(0) = \int_0^0 s\varphi(s)ds + \int_0^1 0\varphi(s)ds = 0$. This means that $0 = \varphi(t) = A\sin(\frac{0}{\sqrt{\lambda}} + B) = A\sin(B)$, so that $B = 0$. From the first derivative, we find that $\varphi'(1) = \int_1^1 \varphi(s)ds = 0$. Then, $0 = \varphi'(1) = \frac{A}{\sqrt{\lambda}}\cos(\frac{t}{\sqrt{\lambda}})$. Since $A = 0$ gives the trivial solution, we instead fix $\frac{1}{\sqrt{\lambda}} = \frac{\pi}{2} + k\pi$, $k \in Z$, so that $\lambda_k = \frac{4}{(2k+1)^2\pi^2}$, for $k \in N$. We then choose A to make this an orthonormal basis, so that $1 = \int_0^1 (A\sin(\frac{2k+1}{2}\pi t))^2 dt = A^2/2$. So $A = \sqrt{2}$. Thus, our basis is $\{\sqrt{2}\sin(\frac{2k+1}{2}\pi t)\}_{k\in N}$, and we may write $W_t = \sum_{k\in N}\sqrt{\lambda_k}\xi_k\varphi_k(t) = \sum_{k\in N}\xi_k\frac{2\sqrt{2}}{(2k+1)\pi}\sin(\frac{2k+1}{2}\pi t)$.

The Karhunen-Loeve Expansion is useful to calculate certain integrals. For example, if $G_t = \sum_{k=0}^{\infty}\sqrt{\lambda_k}\xi_k\varphi_k(t)$ for some $\{(\lambda_k, \varphi_k)\}_{k\in N}$, then we have:

$$
\begin{aligned}
E(\exp(-\frac{\mu}{2}\int_0^1 G_t^2 dt)) &= E(\exp(-\frac{\mu}{2}\int_0^1 (\sum_{k\in N}\sqrt{\lambda_k}\xi_k\varphi_k(t))^2 dt)) \\
&= E(\exp(-\frac{\mu}{2}\int_0^1 (\sum_{k\in N}\sum_{k'\in N}\sqrt{\lambda_k\lambda_{k'}}\xi_k\xi_{k'}\varphi_k(t)\varphi_{k'}(t))dt)) \\
&= E(\exp(-\frac{\mu}{2}\sum_{k\in N}\sum_{k'\in N}(\int_0^1 \sqrt{\lambda_k\lambda_{k'}}\xi_k\xi_{k'}\varphi_k(t)\varphi_{k'}(t)dt))) \\
&= E(\prod_{k\in N}\prod_{k'\in N}\exp(-\frac{\mu}{2}\int_0^1 \sqrt{\lambda_k\lambda_{k'}}\xi_k\xi_{k'}\varphi_k(t)\varphi_{k'}(t)dt)) \\
&= \prod_{k\in N}\prod_{k'\in N}E(\exp(-\frac{\mu}{2}\int_0^1 \sqrt{\lambda_k\lambda_{k'}}\xi_k\xi_{k'}\varphi_k(t)\varphi_{k'}(t)dt)) \\
&= \prod_{k\in N}\prod_{k'\in N}E(\exp(-\frac{\mu}{2}\sqrt{\lambda_k\lambda_{k'}}\xi_k\xi_{k'}\int_0^1 \varphi_k(t)\varphi_{k'}(t)dt)) \\
&= \prod_{k\in N}\prod_{k'\in N}E(\exp(-\frac{\mu}{2}\sqrt{\lambda_k\lambda_{k'}}(\delta_{k,k'}\xi_k^2)(\delta_{k,k'}))) \\
&= \prod_{k\in N}E(\exp(-\frac{\mu}{2}\lambda_k\xi_k^2)) \\
&= \prod_{k\in N}\int_R e^{-\frac{\mu}{2}\lambda_k z^2}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \\
&= \prod_{k\in N}\int_R \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(\mu\lambda_k+1)z^2}dz \\
&= \prod_{k\in N}\frac{1}{\sqrt{1+\mu\lambda_k}}
\end{aligned}
$$

This may be evaluated for any expansion for which this will converge.

4.4 The Weiner Measure

Given a partition of time, $0 = t_0 < t_1 < ... < t_n \leq 1$, we have the probability density function:

$$\begin{aligned}
\rho_{t_n \ldots t_1}(x_n, ..., x_1) &= Z_n^{-1} \exp(-\frac{1}{2} I_n) \\
Z_n &= (2\pi)^{-\frac{n}{2}} \prod_{j=1}^{n} (t_j - t_{j-1})^{-\frac{1}{2}} \\
I_n &= \sum_{j=1}^{n} \frac{(x_j - x_{j-1})^2}{(t_j - t_{j-1})} = \sum_{j=1}^{n} \frac{(x_j - x_{j-1})^2}{(t_j - t_{j-1})^2} (t_j - t_{j-1})
\end{aligned}$$

This can be calculuated for any numerable partition. Since knowing about every numerable partition completely defines a continuous function and the Weiner process is almost always continuous, this gives us information about the entire Weiner process, if we consider the partitions with $t_j - t_{j-1} = \Delta t = \frac{1}{n}$ as $n \to \infty$.

Let $h(t)$ be a function with $X_j = h(\frac{j}{n})$. Then,

$$\begin{aligned}
I_n &= \sum_{j=1}^{n} (\frac{h(\frac{j}{n}) - h(\frac{j-1}{n})}{1/n})^2 (\frac{1}{n}) \\
&\to \int_0^1 (\frac{dh}{dt})^2 dt = I
\end{aligned}$$

This means that the "density" converges to $Z^{-1} \exp(-\frac{1}{2} \int_0^1 (\frac{dh}{dt})^2 dt) Dh(t)$, where $Dh(t) = \prod_{0 \leq t \leq 1} dh_t$. However, Z_n will not converge, so we formally define $Z = \int \exp(-\frac{1}{2} \int_0^1 (\frac{dh}{dt})^2 dt) Dh(t)$. This gives the Weiner measure as

$$d\mu_W = Z^{-1} \exp(-\frac{1}{2} I(h(t))) Dh(t)$$

5 Stochastic Differential Equations

The most general form of a stochastic differential equation is:

$$X_{t_{n+1}}^N = X_{t_n}^N + b(X_{t_n}^N, \{\xi_{t_k}\}_{k \leq n}) \Delta t + \sigma(X_{t_n}^N, \{\xi_{t_k}\}_{k \leq n}) \sqrt{\Delta t} \xi_{t_{n+1}}$$

where b and σ are fixed functions that may depend on random inputs which depend only on past values of the random variable ξ_{t_n}. Under certain conditions, X_t^N will converge is in distribution to X_t, with $\sup_{0 \leq t \leq T} E|X_t - X_t^N| \leq C_1 \sqrt{\Delta t}$ and $\sup_{0 \leq t \leq T} |E(f(E_t)) - E(f(X_t^N))| \leq C_2 \Delta t$, where $\Delta t = \frac{1}{N}$.

The simplest example is the Weiner Process, which has $b = 0$ and $\sigma = 1$. Two other ways of writing this stochastic differential equation are

$$\begin{aligned}
X_{t_{n+1}}^N &= X_{t_n}^N + b(X_{t_n}^N, \{W_{t_k}\}_{k \leq n}) \Delta t + \sigma(X_{t_n}^N, \{W_{t_k}\}_{k \leq n}) \sqrt{\Delta t} (W_{t_{n+1}} - W_{t_n}) \\
dX_t &= b(X_t, W_{[0,t]}) dt + \sigma(X_t, W_{[0,t]}) dW_t
\end{aligned}$$

The latter of these suggests the solution

$$X_t = x_0 + \int_0^t b(X_s, W_{[0,s]})ds + \int_0^t \sigma(X_s, W_{[0,s]})dW_s$$

5.1 Itô Isometries and Formula

Theorem 17 *$E(\int_0^t b(W_s)dW_s) = 0$. $E((\int_0^t b(W_s)dW_s)^2) = \int_0^t E(b(W_s)^2)ds$.*

Proof. At a discrete level, define $I_n = \sum_{j=1}^n b(W_j)(W_{j+1} - W_j)$, where $W_j = W_{t_j}$. Note that I_n converges to $\int_0^t b(W_s)dW_s$. For the first equation, we find that

$$\begin{aligned} E(I_n) &= \sum_{j=1}^n E(b(W_j)(W_{j+1} - W_j)) \\ &= \sum_{j=1}^n E(b(W_j))E(W_{j+1} - W_j) \\ &= \sum_{j=1}^n E(b(W_j))0 \\ &= 0 \end{aligned}$$

using the fact that $b(W_j)$ depends only on past values of W_t and therefore is independent of $W_{j+1} - W_j$. Taking the limit of this, we find that $E(\int_0^t b(W_s)dW_s) = 0$. For the second equation, we find:

$$\begin{aligned} E(I_n^2) &= \sum_{j=1}^n \sum_{k=1}^n E(b(W_j)b(W_k)(W_{j+1} - W_j)(W_{k+1} - W_k)) \\ &= \sum_{j=1}^n E(b(W_j)b(W_j)(W_{j+1} - W_j)(W_{j+1} - W_j)) + \sum_{j=1}^n \sum_{k=1}^{j-1} 0 + \sum_{j=1}^n \sum_{k=J=1}^n 0 \\ &= \sum_{j=1}^n E(b(W_j)^2(W_{j+1} - W_j)^2) \\ &= \sum_{j=1}^n E(b(W_j)^2)E((W_{j+1} - W_j)^2) \\ &= \sum_{j=1}^n E(b(W_j)^2)\Delta t \end{aligned}$$

In the limit, then, we find that $E((\int_0^t b(W_s)dW_s)^2) = \int_0^t E(b(W_s)^2)ds$. ■

Note that this cannot be done if we use other definitions of the integral (such as the midpoint approximation) because we would not have independent intervals.

Theorem 18 *(Itô's formula) If* $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ *and* $Y_t = f(X_t)$, *then* $dY_t = df(X_t) = f'(X_t)(b(X_t)dt + \sigma(X_t)dW_t) + \frac{1}{2}f''(X_t)\sigma(X_t)^2 dt$. *Note that this last term makes this formula different from the standard chain rule.*

Proof. Using the discrete form of the stochastic differential equation (and then taking the limit), we have:

$$X_{n+1} = X_n + b(X_n)\Delta t + \sigma(X_n)(W_{n+1} - W_n)$$

Note that we many consider $W_{n+1} - W_n$ has the product $\sqrt{\Delta t}\xi_{n+1}$ with a standard normal random variable. Using the Taylor expansion for f, up to terms of order $(\Delta t)^1$we find:

$$\begin{aligned} f(X_{n+1}) - f(X_n) &= f(X_n + b(X_n)\Delta t + \sigma(X_n)(W_{n+1} - W_n)) - f(X_n) \\ &= f(X_n) + f'(X_n)(b(X_n)\Delta t + \sigma(X_n)\xi_{n+1}\sqrt{\Delta t}) + \frac{1}{2}f''(X_n)(b(X_n)\Delta t + \sigma(X_n)\xi_{n+1}\sqrt{\Delta t}) \\ &= f'(X_n)(b(X_n)\Delta t + \sigma(X_n)\xi_{n+1}\sqrt{\Delta t}) + \frac{1}{2}f''(X_n)\sigma(X_n)^2\xi_{n+1}^2\Delta t \end{aligned}$$

If we consider ξ_{n+1} as a random variable that takes only the values $+1$ and -1 with probability $1/2$, then $\xi_{n+1}^2 = 1$. Alternately, we may note that $\sum_{j=1}^n \xi_j^2 \Delta t = \sum_{j=1}^n \frac{\xi_j^2}{n}\frac{n}{N}$ which converges to $E(\xi^2)t = Var(\xi)t = t$. Thus, we may say that its derivative with respect to t is 1. Thus,

$$df(X_t) = f'(X_t)(b(X_t)dt + \sigma(X_t)dW_t) + \frac{1}{2}f''(X_t)\sigma(X_t)^2 dt$$

■

Corollary 19 *If we have* $Y_t = g(X_t, t)$ *and* $dX_t = b(X_t)dt + \sigma(X_t)dW_t$, *then* $dg(X_t, t) = \frac{\partial}{\partial X_t}g(X_t, t)dX_t + \frac{1}{2}\frac{\partial^2}{\partial X_t^2}g(X_t, t)\sigma(X_t)^2 dt + \frac{\partial}{\partial t}g(X_t, t)dt$. *(This is the same formula above except for the final time component.)*

Corollary 20 $E((\int_0^t f(W_s)dW_s)(\int_0^{t'} f(W_s)dW_s)) = \int_0^{\min(t,t')} E(f(W_s)^2)ds$. *(This can help find covariances.)*

For example, $df(W_t, t) = (\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial w^2})dt + \frac{\partial f}{\partial w}dW_t$.

In addition, if $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ and $dY_t = c(Y_t)dt + \tau(Y_t)dt$, for the same realization of dW_t, then:

$$d(X_tY_t) = Y_t dX_t + X_t dY_t + \sigma(X_t)\tau(Y_t)dt$$

More generally, we note that $dW_t^2 = dt$ and $\int_0^t dW_s = t$.

5.2 Solving Stochastic Differential Equations

Let $f(t)$ be any function. Given a stochastic differential equation for X_t, define $Y_t = f(t)X_t$. Then,

$$\begin{aligned} dY_t &= d(f(t)X_t) \\ &= f'(t)X_t dt + f(t)dX_t \\ &= (f'(t)X_t + f(t)b(X_t))dt + f(t)\sigma(t)dW_t \end{aligned}$$

In some cases, we may choose $f(t)$ so that the coefficient on dt does not depend on X_t. In that case, we find that $f(t)X_t$ is the sum of a fixed function of t and an integral with respect to dW_t; we may understand the latter by noting that its expected value is 0 and we may calculate its covariance as well using the Ito Isometries.

Let $f(x)$ be any function. Given the stochastic differential equation $dX_t = b(X_t)dt + \sigma(X_t)dW_t$, we may set $Y_t = f(X_t)$ and then find:

$$\begin{aligned} dY_t &= d(f(X_t)) \\ &= f'(X_t)dX_t + \frac{1}{2}f''(X_t)\sigma^2(X_t)dt \\ &= (f'(X_t)b(X_t) + \frac{1}{2}f''(X_t)\sigma^2(X_t))dt + f'(X_t)\sigma(X_t)dW_t \end{aligned}$$

As before, we may be able to choose $f(t)$ so that the coefficient on dt is 0.

5.3 Examples of Stochastic Differential Equations and their Solutions

5.3.1 $X_t = \int_0^t W_s dW_s = \frac{1}{2}W_t^2 - \frac{t}{2}$

We begin by guessing what would be the solution in ordinary differential equations: $Y_t = \frac{1}{2}W_t^2$. We find dY_t:

$$dY_t = \frac{1}{2}d({W_t}^2) = W_t dW_t + \frac{1}{2}dt$$

This implies that $W_t dW_t = \frac{1}{2}d(W_t^2) - \frac{1}{2}dt$, so that

$$\begin{aligned} X_t &= \int_0^t W_s dW_s \\ &= \frac{1}{2}\int_0^t d(W_s^2) - \frac{1}{2}\int_0^t dt \\ &= \frac{1}{2}W_t^2 - \frac{t}{2} \end{aligned}$$

We may check this with the Itô isometries:

$$\begin{aligned}
E(\frac{1}{2}W_t^2 - \frac{t}{2}) &= \frac{1}{2}t - \frac{t}{2} = 0 \\
E((\frac{1}{2}W_t^2 - \frac{t}{2})^2) &= \frac{1}{4}E(W_t^4) - \frac{1}{2}tE(W_t^2) + \frac{1}{4}t^2 \\
&= \frac{1}{4}(3t^2) - \frac{1}{2}t(t) + \frac{1}{4}t^2 \\
&= \frac{1}{2}t^2 \\
&= \int_0^t E(W_s^2)ds
\end{aligned}$$

5.3.2 $dX_t = -\gamma X_t dt + \sigma dW_t$ (Orstein-Uhlenbeck)

Note that

$$\begin{aligned}
d(e^{\gamma t}X_t) &= \gamma e^{\gamma t}X_t dt + e^{\gamma t}dX_t \\
&= \gamma e^{\gamma t}X_t dt + e^{\gamma t}(-\gamma X_t dt + \sigma dW_t) \\
&= \sigma e^{\gamma t}dW_t
\end{aligned}$$

Thus, we know that

$$\begin{aligned}
e^{\gamma t}X_t - x_0 &= \sigma \int_0^t e^{\gamma s}dW_s \\
X_t &= x_0 e^{-\gamma t} + \sigma \int_0^t e^{\gamma s}dW_s
\end{aligned}$$

This allows us to find some properties. First, X_t is Gaussian, since it is the sum of a fixed number and an infinite linear combination of Gaussian random variables. Second, $E(X_t) = x_0 e^{-\gamma t}$, since the second term is 0 by the first Ito Isometry. In addition,

$$\begin{aligned}
E(X_t^2) &= x_0^2 e^{-2\gamma t} + 2x_0\sigma e^{-\gamma t}E(\int_0^t e^{-\gamma(t-s)}dW_s) + \sigma^2 E((\int_0^t e^{-\gamma(t-s)}dW_s)^2) \\
&= x_0^2 e^{-2\gamma t} + 0 + \sigma^2 \int_0^t E(e^{-\gamma(t-s)})^2 ds \\
&= x_0^2 e^{-2\gamma t} + \sigma^2 \int_0^t e^{-2\gamma(t-s)}ds \\
&= x_0^2 e^{-2\gamma t} + \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t})
\end{aligned}$$

Thus, $Var(X_t) = E(X_t^2) - E(X_t)^2 = x_0^2 e^{-2\gamma t} + \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t}) - (x_0 e^{-\gamma t})^2 = \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t})$. Thus, we can consider this as a combination of drift ($x_0 e^{-\gamma t}$)

and a noise term. In addition, we may find the covariance, if $t > s$:

$$\begin{aligned}
E(X_tX_s) &= x_0^2e^{-\gamma(t+s)} + \sigma x_0 e^{-\gamma t}E(\int_0^s e^{-\gamma(s-z)}dW_z) + \sigma x_0 e^{-\gamma s}E(\int_0^t e^{-\gamma(t-z)}dW_z) + \sigma^2 E((\int_0^t e^{-\gamma(t-z)}dW \\
&= x_0^2e^{-\gamma(t+s)} + 0 + 0 + \sigma^2 E((\int_0^s e^{-\gamma(t-z)}dW_z + \int_s^t e^{-\gamma(t-z)}dW_z)(\int_0^s e^{-\gamma(s-z)}dW_z)) \\
&= x_0^2e^{-\gamma(t+s)} + \sigma^2 E((\int_0^s e^{-\gamma(t-z)}dW_z)(\int_0^s e^{-\gamma(s-z)}dW_z)) + \sigma^2 E((\int_s^t e^{-\gamma(t-z)}dW_z)(\int_0^s e^{-\gamma(s-z)} \\
&= x_0^2e^{-\gamma(t+s)} + \sigma^2 e^{-\gamma t}e^{-\gamma s}E((\int_0^s e^{\gamma z}dW_z)^2) + \sigma^2 E((\int_s^t e^{-\gamma(t-z)}dW_z)(\int_0^s e^{-\gamma(s-z)}dW_z)) \\
&= x_0^2e^{-\gamma(t+s)} + \sigma^2 e^{-\gamma t}e^{-\gamma s}\int_0^s E(e^{2\gamma z})dz + 0 \\
&= x_0^2e^{-\gamma(t+s)} + \sigma^2 e^{-\gamma(t+s)}\frac{1}{2\gamma}(e^{2\gamma} - 1)
\end{aligned}$$

In the limit, with $|t-s|$ is fixed, then the covariance is $\frac{\sigma^2}{2\gamma}e^{-\gamma|t-s|}$.

5.3.3 $dX_t = -\gamma X_t dt + \sigma X_t dW_t$

Note that this is equivalent to $\frac{1}{X_t}dX_t = -\gamma dt + \sigma dW_t$. We consider $g(X_t) = \ln(X_t)$. Using Ito's formula, we find that

$$\begin{aligned}
d(\ln X_t) &= \frac{1}{X_t}dX_t - \frac{1}{2X_t^2}\sigma^2 X_t dt \\
&= \frac{1}{X_t}(-\gamma X_t dt + \sigma X_t dW_t) - \frac{1}{2X_t^2}\sigma^2 X_t^2 dt \\
&= -\gamma dt + \sigma dW_t - \frac{\sigma^2}{2}dt
\end{aligned}$$

Integrating, we find that:

$$\begin{aligned}
\ln X_t - \ln x_o &= -\gamma t + \sigma W_t - \frac{\sigma^2}{2}t \\
\ln(\frac{X_t}{x_0}) &= -(\gamma + \frac{\sigma^2}{2})t + \sigma W_t \\
X_t &= x_0 e^{-(\gamma + \frac{\sigma^2}{2})t + \sigma W_t}
\end{aligned}$$

Notice that $X_t^n = x_0^n \exp(-n(\gamma + \frac{\sigma^2}{2})t + n\sigma W_t)$. Then,

$$\begin{aligned}
E(X_t^n) &= x_0^n E(\exp(-n(\gamma + \frac{\sigma^2}{2})t + n\sigma W_t)) \\
&= x_0^n \exp(-n(\gamma + \frac{\sigma^2}{2})t)E(e^{n\sigma W_t}) \\
&= x_0^n \exp(-n(\gamma + \frac{\sigma^2}{2})t)\exp(\frac{1}{2}n^2\sigma^2 t)
\end{aligned}$$

$dX_t = \sigma X_t dW_t$ We know from above that $X_t = \exp(-\frac{\sigma^2}{2}t + \sigma W_t) = f(\sigma)$. We could also consider this differential equation iteratively:

$$\begin{aligned} X_t &= 1 + \sigma \int_0^t X_s dW_s \\ &= 1 + \sigma(1 + \int_0^t (1 + \sigma \int_0^s X_u dW_u) dW_s) \\ &= ... \\ &= 1 + \sum_{n=1}^{\infty} \sigma^n \int ... \int_{0 \leq s_1 \leq ... \leq s_n \leq t} dW_{s_n} ... dW_{s_1} \end{aligned}$$

Since we can also write a Taylor expansion of $f(\sigma) = \exp(-\frac{\sigma^2}{2}t + \sigma W_t)$, we could use the coefficients to compute the values of $\int ... \int_{0 \leq s_1 \leq ... \leq s_n \leq t} dW_{s_n} ... dW_{s_1}$ for any n.

5.4 Approximation Schemes for stochastic differential equations

We may write $X_{t+\Delta t} = X_t + \int_t^{t+\Delta t} b(X_s)ds + \int_t^{t+\Delta t} \sigma(X_s)dW_s$. A first approximation of this is $\widehat{X_{t+\Delta t}} = X_t + b(\widehat{X_t})\Delta t + \sigma(\widehat{X_t})(W_{t+\Delta t} - W_t)$. As with other stochastic approximations, this must be evaluated at the beginning of each interval, not at intermediate points. We evaluate this approximation by noting that $\sup_{0 \leq t \leq T} E(|X_t - \widehat{X_t}|) \leq C\sqrt{\Delta t}$ (and the approximation is of strong order $\frac{1}{2}$) and $\sup_{0 \leq t \leq T} |E(f(X_t)) - E(f(\widehat{X_t}))| \leq C\Delta t$ for suitable test functions f (and the approximation is of weak order 1). To improve this, we may use higher order terms as well. One such scheme is:

$$\widehat{X_{t+\Delta t}} = X_t + b(\widehat{X_t})\Delta t + \sigma(\widehat{X_t})(W_{t+\Delta t} - W_t) - \frac{1}{2} b(\widehat{X_t}) b'(\widehat{X_t})((W_{t+\Delta t} - W_t)^2 - \Delta t)$$

This is called the Milstein (Talay) Approximation.

6 Path Integral Representations of Stochastic Differential Equations

Given a one-dimensional random variable, Z, with probability density function, $\rho(z)$, and a new random variable $X = \Phi(Z)$, then we may write:

$$\begin{aligned} E(f(X)) &= \int_R f(\Phi(Z))\rho(z)dz \\ &= \int_R f(x)\rho(\Phi^{-1}(x))\frac{dz}{dx}dx \\ &= \int_R f(x)\widehat{\rho}(x)dx \end{aligned}$$

where $\widehat{\rho}$ is a pdf for X. In addition, if $\widehat{\rho}(x) = Z^{-1}\exp(-\frac{1}{2}x^2 + g(x))$, where Z is the normalizing constant:

$$\begin{aligned} E(f(X)) &= \frac{\int_R f(x)\exp(-\frac{1}{2}x^2 + g(x))dx}{\int_R \exp(-\frac{1}{2}x^2 + g(x))dx} \\ &= \frac{\frac{\int_R f(x)\exp(-\frac{1}{2}x^2+g(x))dx}{\int_R \exp(-\frac{1}{2}x^2)dx}}{\frac{\int_R \exp(-\frac{1}{2}x^2+g(x))dx}{\int_R \exp(-\frac{1}{2}x^2)dx}} \\ &= \frac{E(f(W)\exp(g(W)))}{E(\exp(g(W)))} \end{aligned}$$

where W is a standard normal random variable.

In the case of a discrete stochastic process $X_{[0.T]} = X(W_{[0,T]})$ with differential equation $X_{n+1} = X_n + b(X_n)\Delta t + (W_{n+1} - W_n)$ evaluated at N intervals of Δt, this becomes:

$$E(f(X)) = \int_{R^N} f(x(w))\rho(w)dw$$

where $\rho(w) = \frac{1}{(2\pi\Delta t)^{N/2}}\exp(-\frac{1}{2\Delta t}\sum_{j=1}^{N}(w_j - w_{j-1})^2)$. Since $X_{n+1} - X_n - b(X_n)\Delta t = (W_{n+1} - W_n)$,

$$\rho(w(x)) = \frac{1}{(2\pi\Delta t)^{N/2}}\exp(-\frac{1}{2}\sum_{j=1}^{N}\frac{(X_{n+1} - X_n - b(X_n)\Delta t)^2}{\Delta t})$$

We find the Jacobian of the transformation by noting that $\frac{\partial X_n}{\partial W_m} = 0$ when $m > n$, since X_n does not depend on future value of W_m. In addition, $\frac{\partial X_n}{\partial W_n} = 1$, since $X_{n+1} = X_n + b(X_n)\Delta t + (W_{n+1} - W_n)$ and the only occurrence of W_{n+1} is in the last term, with a coefficient of 1. (Note that this step relies on the assumption that $\sigma(X_n) = 1$.) Thus, $[\frac{\partial X_n}{\partial W_m}]$ is a triangular matrix with only ones on the diagonal, and the Jacobian is 1. That means that

$$\begin{aligned} \widehat{\rho}(x) &= \frac{1}{(2\pi\Delta t)^{N/2}}\exp(-\frac{1}{2}\sum_{j=1}^{N}\frac{(X_{n+1} - X_n - b(X_n)\Delta t)^2}{\Delta t}) \\ E(f(X)) &= \int_{R^N} f(x)\widehat{\rho}(x)dx_1...dx_N \end{aligned}$$

This gives a mapping from the density of paths of $W_{[0,T]}$ to the density of paths in $X_{[0,T]}$.

In addition, this gives us another expression for $E(f(X))$:

$$\begin{aligned} E(f(X)) &= \frac{\int_{R^N} f(x(w))\exp(-\frac{1}{2}\sum_{j=1}^{N}\frac{(w_j - w_{j-1})^2}{\Delta t})dw_1...dw_N}{\int_{R^N}\exp(-\frac{1}{2}\sum_{j=1}^{N}\frac{(w_j - w_{j-1})^2}{\Delta t})dw_1...dw_N} \\ &= \frac{\int_{R^N} f(x)\exp(-\frac{1}{2}\sum_{j=1}^{N}\frac{(x_j - x_{j-1} - b(x_{j-1}))^2}{\Delta t})dx_1...dx_N}{\int_{R^N}\exp(-\frac{1}{2}\sum_{j=1}^{N}\frac{(x_j - x_{j-1} - b(x_{j-1}))^2}{\Delta t})dx_1...dx_N} \end{aligned}$$

Extending to the continuous case (and a path integral), we note that:

$$E(f(X)) = \frac{\int f(h_{[0,T]}) \exp(-\frac{1}{2}\int_0^T (h'(t) - b(h(t)))dt) dH_{[0,T]}}{\int \exp(-\frac{1}{2}\int_0^T (h'(t) - b(h(t)))dt) dH_{[0,T]}}$$

This allows us to compute expectations for the stochastic process defined by $dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t$ with $X_0 = x$ in terms of expectations of the Weiner process:

$$E(f(X_{[0,T]})) = \frac{E(f(x + W_{[0,T]}) \exp(\int_0^T b(x + W_{[0,T]})dW_t - \frac{1}{2}\int_0^T b(x + W_t)^2 dt))}{E(\exp(\int_0^T b(x + W_{[0,T]})dW_t - \frac{1}{2}\int_0^T b(x + W_t)^2 dt))}$$

We show that this denominator is 1. Let $Q_t = \int_0^t b(W^x_{[0,s]})dW_s - \frac{1}{2}\int_0^t b(W^x_{[0,s]})^2 ds$. Then:

$$\begin{aligned}
dQ_t &= b(W^x_t)dW_t - \frac{1}{2}b(W^x_t)^2 dt \\
dZ_t &= d(e^{Q_t}) \\
&= e^{Q_t}(b(W^x_t)dW_t - \frac{1}{2}b(W^x_t)^2 dt) + \frac{1}{2}e^{Q_t}b(W^x_t)^2 dt \\
&= Z_t b(W^x_t)dW_t
\end{aligned}$$

Since $Q_0 = 0$, $Z_0 = e^0 = 1$, and $Z_t = 1 + \int_0^t Z_s b(W^x_s)dW_s$. Using the first Ito Isometry, $E(Z_t) = 1 + E(\int_0^t Z_s b(W^x_s)dW_s) = 1$. Returning to the formula for expectations of functionals of a stochastic process:

$$E(f(X_{[0,T]})) = E(f(W^x_{[0,T]}) \exp(\int_0^T b(W^x_t)dW_t - \frac{1}{2}\int_0^T b(W^x_t)^2 dt))$$

As an example, if we have $dX_t = -X_t dt + dW_t$ and $f(X_t) = \exp(-\alpha \int_0^T X_t^2 dt)$, then,

$$\begin{aligned}
E(\exp(-\alpha \int_0^T X_t^2 dt)) &= E(f(X_{[0,T]})) \\
&= \frac{E(\exp(-\alpha \int_0^T W_t^2 dt) \exp(\int_0^T W_t dt - \frac{1}{2}\int_0^T W_t^2 dt))}{E(\exp(\int_0^T W_t dt - \frac{1}{2}\int_0^T W_t^2 dt))}
\end{aligned}$$

6.1 The Girsanov Principle

We may generalize to relationships between other stochastic processes. Suppose $dX_t = b(X_t)dt + dW_t$ and $dY_t = c(X_t)dt + dW_t$. Then, discretizing, squaring, and substituting, we find

$$E(f(X_{[0,T]})) = E(f(Y_{[0,T]}) \exp(\int_0^T (b(Y_t) - c(Y_t))dY_t - \int_0^T c(Y_t)(b(Y_t) - c(Y_t))dt + \frac{1}{2}\int_0^T (b(Y_t) - c(Y_t))^2 dt))$$

This can help prove the existence of related processes. In addition, this is a convenient way to remove or change the drift term.

6.1.1 An Example using Girsanov

Let $dX_t = b(t, X_t)dt + dW_t,\ X_0 = 0$. Let $dP(w)$ be the Weiner measure. Define $M_t = \exp(-\int_0^t b(t, X_t)dW_s - \frac{1}{2}\int_0^t b(t, X_t)ds)$. Define a new measure by $dQ(w) = M_T dP(w)$. With respect to this new measure, X_t is Brownian motion; that is, $dX_t = d\widetilde{W_t}$ with respect to the measure Q.

6.2 Theory of Large Deviations

According to the Laplace method, if a function $g(x)$ has a unique minimum at x_0, then $-\varepsilon \ln \int_R f(x)e^{-\frac{1}{\varepsilon}g(x)}dx \to g(x_0)$ as $\varepsilon \to 0$ for any reasonable function $f(x)$. We say that $\int_R f(x)e^{-\frac{1}{\varepsilon}g(x)}dx \asymp e^{-\frac{1}{\varepsilon}g(x_0)}$ (this is asymptotic equality). Extending this to the continuous case, we find that

$$\int f(h_{[0,T]})\exp(-\frac{1}{2\varepsilon}[\int_0^T (h'(t) - b(h(t)))^2 dt]Dh_{[0,T]}) \asymp \exp(-\frac{1}{2\varepsilon}[\int_0^T (h_0'(t) - b(h_0(t)))^2 dt]Dh_{0[0,T]})$$

where h_0 minimizes $\int_0^T (h'(t) - b(h(t)))^2 dt$. If there are no constraints, the minimizer is simply the solution to the ordinary differential equation $h'(t) = b(h(t))$, in which case $g(h_{0[0,T]}) = \int_0^T 0dt = 0$.

Let $dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}dW_t$, for $0 < \varepsilon << 1$. As $\varepsilon \to 0$, $X_{[0,T]}^\varepsilon \to X_{[0,T]}$, where $dX_t = b(X_t)$. We may also find the probability of "rare events" in which X_t^ε deviates greatly from its expected path. Let $\varphi(t)$be any deterministic function. Let $f(X_{[0,t]}) = \int_0^T (X_t^\varepsilon - \varphi(t))^2 dt$; then $E(f(X_{[0,T]}^\varepsilon))$ can give us an idea of the deviation. Let $M[h] = \exp(-\frac{1}{2\varepsilon^2}\int_0^T (h'(t) - b(h(t)))^2 dt)$; this is proportional to the density of X_t (or something). Given any functional $F[h]$, we know that

$$\begin{aligned} E(F(X_t)) &= \frac{\int F[h]M[h]dh}{\int M[h]dh} \\ P(X_T &> a) = \frac{\int_{h(T)>a} M[h]dh}{\int M[h]dh} \end{aligned}$$

We evaluate the numerator of second expression by minimizing $\int_0^T (h'(t) - b(h(t)))^2 dt$ subject to the constraint that $h(T) > a$. To do this, we note that we may apply the Laplace method and minimize $\int_0^T (h'(t) - b(h(t)))^2 dt$ subject to the relevant constraints. Integrating by parts, we find that we must solve:

$$0 = (h_0' + b(h_0))' + (h_0' + b(h_0))b'(h_0)$$

Let $p(t) = (h_0' + b(h_0))$. Then we must solve $h_0' = b(h_0) + p$ and $p' = -b'(h_0)p$ subject to the boundary conditions; note that this looks like the stochastic differential equation with the stochastic term $\sqrt{\varepsilon}dW_t$ replaced by $p(t)$.

6.2.1 An Example

Let $dX_t^\varepsilon = -X_t^\varepsilon dt + \varepsilon dW_t$, with $X_0 = x$. The deterministic solution of this differential equation is $X_t = xe^{-t}$. In this case, $M[h] = \exp(-\frac{1}{2\varepsilon^2}\int_0^T (h'(t) + h(t))^2 dt)$, and we must minimize $\int_0^T (h'(t) + h(t))^2 dt$ subject to the constraints that $h(0) = x$ and $h(T) > a$. Suppose h_0 minimizes the integral subject to the constraints. Then, for any arbitrary function h_1,

$$\int_0^T (h_0'+h_1'+h_0+h_1)^2 dt = \int_0^T (h_0'+h_0)^2 dt + 2\int_0^T (h_0'+h_0)(h_1'+h_1)dt + \int_0^T (h_1'+h_1)^2 dt$$

Note that the last term must be positive. We set the middle term to 0 and integrate by parts to find:

$$\begin{aligned} 0 &= \int_0^T (h_0' + h_0)(h_1' + h_1)dt \\ &= \int_0^T (-(h_0' + h_0)' h_1 + (h_0' + h_0)h_1)dt \end{aligned}$$

Since h_0 and $h_0 + h_1$ satisfy the boundary conditions $h(0) = x$ and $h(T) > a$, $h_1(0) = h_1(T) = 0$, which is how the integration by parts worked. Thus, we find a minimum when $(h_0' + h_0)' = (h_0' + h_0)$, that is, $h_0'' = h_0$. The solution is $h(t) = Ae^{-t} + Be^t$, with $x = h(0) = A + B$ and $a = h(T) = Ae^{-T} + Be^T$. Thus,

$$\begin{aligned} M[h_0] &= \exp(-\frac{1}{2\varepsilon^2}\int_0^T 2(a\frac{e^t}{e^T - e^{-T}})^2 dt) \\ &= \exp(-\frac{1}{\varepsilon^2}\int_0^T \frac{a^2}{(e^T - e^{-T})^2} e^{2t} dt) \\ &= \exp(-\frac{a^2}{\varepsilon^2}\frac{e^{2T} - 1}{(e^T - e^{-T})^2}) \end{aligned}$$

By the Laplace method, $P(X_T > a) = \int_{h(T)>a} M[h]dh \asymp \exp(-\frac{a^2}{\varepsilon^2}\frac{e^{2T}-1}{(e^T-e^{-T})^2})$. (Note that any path that does go above a is likely to look similar to h_0.)

6.2.2 Applying the Girsanov Principle

We may also use the Girsanov principle, using $dX_t = b(X_t)dt + \varepsilon dW_t$ and $dY_t = \phi(t)dt + \varepsilon dW_t$. note that

$$\begin{aligned} (h' - b(h))^2 &= (h' - \phi + \phi - b(h))^2 \\ &= (h' - \phi)^2 + 2(h' - \phi)(\phi - b(h)) + (\phi - b(h))^2 \end{aligned}$$

This means that the density of Y_t is proportional to $\exp(-\frac{1}{2\varepsilon^2}\int_0^T ((h' - \phi)^2 + 2(h' - \phi)(\phi - b(h)) + (\phi - b(h))^2)dt)$, and the first term provides the weights for Y. Note that $(h(t) - \phi(t)) = \frac{1}{\varepsilon}(dY_t - \phi(t)) = dW_t$. Thus, $M(Y_{[0,T]}) =$

$\exp(\frac{1}{\varepsilon}\int_0^T (b(Y_t)-\phi(t))dW_t - -\frac{1}{2\varepsilon^2}\int_0^T (b(Y_t)-\phi(t))^2 dt)$. Returning to the probability we are calculating, we recall that:

$$P(X_T \geq a) = E(1_{X_T \geq a}) = E(1_{Y_T \geq a} M(Y_{[0,T]}))$$

This formula does not depend on ϕ, so we may choose $\phi(t) = h_0'(t)$, so that it solves $0 = (h_0' - b(h_0))' + (h_0' - b(h_0)b'(h_0)$. Then, $Y_t = h_0(t) + \varepsilon W_t$ (since we may just integrate the two terms separately). Substituting for Y_t, we find:

$$M(Y_{[0,T]}) = \exp(\frac{1}{\varepsilon}\int_0^T (b(h_0(t)+\varepsilon W_t) - h_0'(t))dW_t - \frac{1}{2\varepsilon^2}\int_0^T (b(h_0(t)+\varepsilon W_t) - h_0'(t))^2 dt)$$

Using a Taylor expansion in ε, the first term of the exponent becomes:

$$\begin{aligned} & \frac{1}{\varepsilon}\int_0^T (b(h_0(t)) + \varepsilon W_t b'(h_0(t)) + O(\varepsilon^2) - h_0'(t))dW_t \\ = & \int_0^T \frac{1}{\varepsilon}(b(h_0(t)) - h_0'(t)) + W_t b'(h_0(t)) + O(\varepsilon))dW_t \end{aligned}$$

and the second term of the exponent becomes:

$$-\frac{1}{2\varepsilon^2}\int_0^T (b(h_0(t)) + \varepsilon W_t b'(h_0(t)) + \frac{1}{2}b''(h_0(t))\varepsilon^2 W_t^2 + O(\varepsilon^3) - h_0'(t))^2 dt$$

The highest order term is $-\frac{1}{2}\int_0^T \frac{1}{\varepsilon^2}(b(h_0(t)) - h_0'(t))^2 dt$, which depends only on non-stochastic functions. We show that the terms with order $\frac{1}{\varepsilon}$ cancel, using integration by parts and the definition of h_0:

$$\begin{aligned} & \frac{1}{\varepsilon}\int_0^T (b(h_0(t)) - h_0'(t))dW_t - \frac{1}{2\varepsilon^2}\int_0^T 2(b(h_0(t)) - h_0'(t))\varepsilon b'(h_0(t))W_t dt \\ = & \frac{1}{\varepsilon}(\int_0^T (b(h_0(t)) - h_0'(t))' W_t dt) - \int_0^T (b(h_0(t)) - h_0'(t)) b'(h_0(t))W_t dt \\ = & \frac{1}{\varepsilon}\int_0^T ((b(h_0(t)) - h_0'(t))' - (b(h_0(t)) - h_0'(t))b'(h_0(t)))W_t dt \\ = & \frac{1}{\varepsilon}\int_0^T 0 W_t dt = 0 \end{aligned}$$

Thus, as $\varepsilon \to 0$,

$$M(Y_{[0,T]}) = \exp(-\frac{1}{2}\int_0^T \frac{1}{\varepsilon^2}(b(h_0(t)) - h_0'(t))^2 dt)\widetilde{M}(Y_{[0,T]}))$$

where $\widetilde{M}(Y_{[0,T]})$ has a limit as $\varepsilon \to 0$. By the Laplace Method,

$$\begin{aligned} -\varepsilon^2 \log P(X_T & \geq a) = -\varepsilon^2 \log(E(1_{Y_T \geq a}\widetilde{M}(Y_{[0,T]})) + \frac{\varepsilon^2}{2\varepsilon^2}\int_0^T (b(h_0(t)) - h_0'(t))^2 dt) \\ -\varepsilon^2 \log P(X_T & \geq a) \to \frac{1}{2}\int_0^T (b(h_0(t)) - h_0'(t))^2 dt \end{aligned}$$

Using lower order terms, we may also find that

$$\lim_{\varepsilon\to 0} \exp(\frac{1}{2\varepsilon^2}\int_0^T (b(h_0(t)) - h_0'(t))^2 dt) P(X_T \geq a) = E(1_{Y_T\geq a}\overline{M}(h_0))$$

where $\overline{M}(h_0) = \exp(\int_0^T b'(h_0(t))W_t dt - \frac{1}{2}\int_0^T((b'(h_0(t)W_t)^2 + 2(b(h_0(t)) - h_0'(t))b''(h_0(t))W_t^2)dt$.

7 The Fokker-Planck Equations

These are also known as the Forward and Backward Kolmogorov equations.

Let $dX_t = b(X_t)dt + \sigma(X_t)dW_t$, with $X_0 = x$. Define $p_t^x(y)$ by $\int_{[a,b]} p_t^x(y)dy = P(X_t \in [a,b])$; that is, $p_t^x(y)$ is the probability density function of X_t for a fixed t. In addition, define $u(x,t) = E(f(X_t^x))$. Note that:

$$u(x,t) = E(f(X_t^x)) = \int_R f(y)p_t^x(y)dy$$

7.1 The Forward Kolmogorov Equation

Using this fact, we find $p_t^x(y)$:

$$\begin{aligned}
df(X_t) &= f'(X_t)(b(X_t)dt + \sigma(X_t)dW_t) + \frac{1}{2}f''(X_t)\sigma^2(X_t)dt \\
f(X_T) &= f(x) + \int_0^T f'(X_t)b(X_t)dt + \int_0^T f'(X_t)\sigma(X_t)dW_t + \int_0^T \frac{1}{2}f''(X_t)\sigma^2(X_t)dt \\
u(x,T) &= E(f(X_T)) \\
&= f(x) + E(\int_0^T f'(X_t)b(X_t)dt) + 0 + E(\int_0^T \frac{1}{2}f''(X_t)\sigma^2(X_t)dt)
\end{aligned}$$

We may write these expectations as integrals using $p_t^x(y)$. Changing the order of integration and then integrating by parts, we find:

$$\begin{aligned}
\int_R f(y)p_T^x(y)dy &= f(x) + \int_0^T \int_R (f'(y)b(X_t) + \frac{1}{2}f''(X_t)\sigma^2(X_t))dydt \\
&= f(x) + \int_0^T \int_R f(y)(-\frac{\partial}{\partial y}(b(y)p_t^x(y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p_t^x(y)))dydt
\end{aligned}$$

Taking the derivative with respect to time yields:

$$\int_R f(y)\frac{\partial}{\partial t}p_t^x(y)dy = \int_R f(y)(-\frac{\partial}{\partial y}(b(y)p_t^x(y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p_t^x(y)))dy$$

Since this is true for any function $f(y)$, we have the Forward Kolmogorov Equation:

$$\frac{\partial}{\partial t}p_t^x(y) = -\frac{\partial}{\partial y}(b(y)p_t^x(y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p_t^x(y))$$

In addition, we have the boundary conditions that $\lim_{t\to 0^+} p_t^x(y) = \delta(x-y)$, since $X_0 = x$, and $\lim_{y\to\infty} \frac{\partial}{\partial y} p_t^x(y) = 0$, because this is a pdf. We will later define:

$$A^*(f(y)) = -\frac{\partial}{\partial y}(b(y)f(y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)f(y))$$

so that we have the equation $\frac{\partial}{\partial t} p_t^x(y) = A^*(p_t^x(y))$.

There is a "statistical steady state" – that is, a limiting distribution – if $p_t^x(y)$ has a limit, in which case there is a finite solution to

$$\begin{aligned}
0 &= -\frac{\partial}{\partial y}(b(y)p_t^x(y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p_t^x(y)) \\
0 &= -b(y)p_t^x(y) + \frac{1}{2}\frac{\partial}{\partial y}(\sigma^2(y)p_t^x(y)) \\
b(y)p_t^x(y) &= \frac{1}{2}\frac{\partial}{\partial y}(\sigma^2(y)p_t^x(y)) \\
\frac{1}{2}\frac{\partial}{\partial y}(\sigma^2(y)p_t^x(y))\frac{1}{p_t^x(y)} &= \frac{b(y)}{\sigma^2(y)} \\
\ln(\sigma^2(y)p_t^x(y)) &= C + \int \frac{b(y)}{\sigma^2(y)} dy \\
p_t^x(y) &= \frac{C}{\sigma^2(y)} \exp(\int \frac{b(y)}{\sigma^2(y)} dy)
\end{aligned}$$

7.2 The Backward Kolmogorov Equation

We will consider $\lim_{t\to 0} \frac{E(f(X_t^x)-f(x))}{t}$. Note that $E(f(X_t^x)-f(x)) = E(\int_0^t f'(X_s)b(X_s)ds + \int_0^t \frac{1}{2} f''(X_s)\sigma^2(X_s)ds)$. Since, recursively, $X_s = x + \int_0^s b(X_z^x)dz + \int_0^s \sigma(X_z^x)dW_z$, $E(f(X_t^x) - f(x)) = t(f'(x)b(x) + \frac{1}{2}f''(x)\sigma^2(x)) + o(t)$. Define the infinitesimal generator, A, on a function f by:

$$A(f(x)) = (b(x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2})f(x)$$

(The A^* defined above satisfies $\int_R g(x)A(f(x))dx = \int_R f(x)A^*(g(x))dx$ for any functions f and g.)

Since $u(x,t) = E(f(X_t^x))$, $\lim_{t\to 0} \frac{\partial u}{\partial t} = A(f(x))$, and $u(x,0) = f(x)$. We will show that $\frac{\partial u}{\partial t} = A(u) = b(x)\frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial u}{\partial x}$ for all t. For simiplicity, we assume that not only is X_t a Markov process, but it is stationary (the second assumption is not quite necessary). Then, $p_{t+h}^x(y) = \int_R p_t^x(z)p_h^z(y)dz$. We calculate:

$$\begin{aligned}
E(u(X_h,t)) &= \int_R p_h^x(y)u(y,t)dy \\
&= \int_{R^2} p_h^x(y)p_t^y(z)f(z)dzdy \\
&= \int_{R^2} p_t^y(z)p_h^x(y)dyf(z)dz \\
&= \int_R p_{t+h}^x(z)f(z)dz \\
&= E(f(X_{t+h}^x)) \\
&= u(x,t+h)
\end{aligned}$$

Thus, $\frac{E(\varphi(X_h^x))-\varphi(x)}{h} = A(\varphi(x))$ for any φ and any t. In particular, $\frac{\partial u}{\partial t} = \frac{E(u(X_t,h))-u(x,t)}{h} = A(u(x,t))$ for all t. Consider the function $f(y) = \delta(y-z)$ for some z. Then, $u(x,t) = \int_R p_t^x(y)f(y)dy = p_t^x(z)$. Using this function in the formula above, we find the Backward Kolmogorov equation:

$$\frac{\partial}{\partial t}p_t^x(y) = b(x)\frac{\partial}{\partial x}p_t^x(y) + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}p_t^x(y)$$

subject to the boundary condition that $\lim_{t\to 0^+} p_t^x(y) = \delta(x-y)$.

The Kolmogorov equations allow us to relate partial differential equations and stochastic differential equations.

7.3 First Passage Time

Given a process X_t with an initial condition x, $a < x < b$, define τ as the first time that X_t moves outside $[a,b]$, that is, $\tau = \inf\{t : X_t \notin [a,b]\}$. Let $m(t)$ be the pdf of τ, so that $\int_0^t m(z)dz$ is the probability that X_t leaves the interval before time t. Just as with discrete time Markov chains, we may modify this chain to set $p_t^x(a) = p_t^x(b) = 0$. This gives additional boundary conditions for the solution of the Forward equation:

$$\frac{\partial}{\partial t}p_t^x(y) = -\frac{\partial}{\partial y}(b(y)p_t^x(y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)p_t^x(y))$$

Using these conditions, we find that

$$\frac{\partial}{\partial t}\int_a^b p_t^x(y)dy = (-b(y)p_t^x(y) + \frac{1}{2}\frac{\partial}{\partial y}(\sigma^2(y)p_t^x(y))_a^b < 0$$

(We would expect that the probability of escape increases with time.) Using the Backward equation, with the same boundary conditions we find that:

$$\begin{aligned}
\frac{\partial}{\partial t}\int_a^b p_t^x(y)dy &= b(x)\frac{\partial}{\partial x}\int_a^b p_t^x(y)dy + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}\int_a^b p_t^x(y)dy \\
\frac{\partial}{\partial t}m^x(t) &= b(x)\frac{\partial m^x(t)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 m^x(t)}{\partial x^2}
\end{aligned}$$

Let $T(x) = E(\tau^x) = \int_0^\infty t m^x(t)dt$. Using the equation above, we find that:

$$\begin{aligned}\int_0^t t\frac{\partial}{\partial t}m^x(t)dt &= b(x)\frac{\partial}{\partial x}\int_0^t m^x(t)dt + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}\int_0^t m^x(t)dt \\ \int_0^\infty m^x(t)dt + (tm^x(t))_0^\infty &= b(x)\frac{\partial}{\partial x}T(x) + \frac{1}{2}\sigma^2(x)\frac{\partial^2}{\partial x^2}T(x)\end{aligned}$$

Since the left-hand-side is the sum of a density and something that is 0 (allegedly), we have the differential equation $1 = b(x)T'(x) + \frac{1}{2}\sigma^2(x)T''(x)$, with the boundary conditions $T(a) = T(b) = 0$, which can be solved with standard methods.