Stochastic Calculus: NYU, Fall 2003

# 1 Preliminaries

## 1.1 Multivariate Normals

**Definition 1** *A Gaussian vector (multivariate normal random variable),* $X = \begin{pmatrix} X_1 \\ \dots \\ X_p \end{pmatrix}$*, 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 $\Sigma$ 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* $\xi_j$ *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,* $P(\left|\frac{S_n}{n} - \eta\right| \leq \varepsilon) \longrightarrow 1$ *as* $n \longrightarrow \infty$ *for all* $\varepsilon > 0$*.*

**Theorem 3** *(Central Limit Theorem) Let* $\xi_j$ *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* $\rho(x) = e^{-x^2/2}/\sqrt{2\pi}$*.*

**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* $\sum_{j=1}^\infty P(B_j) = \infty$*.*

**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* $E(|X_j|) < \infty$*.*

## 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.*

Some properties of Markov chains with a state space, $S$, initial distribution, $\mu$, 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 $\pi$ such that $\pi = P\pi$ and (2) for all $\mu_0$ then $\mu_n = P^n\mu_0$ converges to $\pi$ as $n \longrightarrow \infty$ exponentially fast.*

**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, $P(W_t \le x | \{W_{s'}\}_{s' \le s}) = P(W_t \le x | W_s)$.
- $P(W_t \le x | W_s = y) = P(W_{t-s} \le x - y)$, and $W_t - W_s$ has the same distribution as $W_{t-s}$, that is, $Normal(0, t-s)$
- $\rho_{t_n - t_{n-1}}(x_n | x_{n-1}) ... \rho_{t_2 - t_1}(x_2 | x_1) \rho_{t_1 - 0}(x_1 | 0) = \rho_{t_n - t_{n-1}}(x_n - x_{n-1} | 0) ... \rho_{t_2 - t_1}(x_2 - x_1 | 0) \rho_{t_1 - 0}(x_1 | 0)$ (this is the the joint probability density function of $W_{t_n}, ..., W_{t_0}$ for any partition $\{t_0 = 0, t_1, ..., t_{n-1}, t_n = 1\}$ of $[0,1]$
- $E(W_s W_t) = \min(s,t)$
- $E((W_t - W_s)^2) = |t - s|$ (The Weiner process is almost surely continuous, and it is almost surely not differentiable anywhere.
- Given $\lambda > 0$, $\lambda^{-1/2} W_{\lambda t}$ is equal in distribution to $W_t$. (The Weiner Process is self-similar.)

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,

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

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, $E(W_tW_s) = \min(t,s)$.

## 4.2 Gaussian Processes

**Definition 12** *$X_t$, 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 $E(X_tX_s) = K(t,s)$ 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 13** *$W_t$ is the Weiner process (Brownian motion) if (1) $W_t$ is a Gaussian process, (2) $E(W_t) = 0$ and $E(W_sW_t) = \min(s,t)$, 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 $k$. 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.

To find a Karhunen-Loeve expansion, solve $\lambda\varphi(t) = \int_0^1 E(B_sB_t)\varphi(s)ds$, subject to the boundary conditions implied by the original equation, such as $\varphi(0) = 0$, and the fact that $\{\varphi_k\}_{k\in N}$ should be orthonormal.

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:

$$E(\exp(-\frac{\mu}{2}\int_0^1 G_t^2dt)) = \prod_{k\in N}\frac{1}{\sqrt{1+\mu\lambda_k}}$$

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

# 5 Stochastic Differential Equations

The most general form of a stochastic differential equation is:

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

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

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

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

**Theorem 15** *(Itô's formula) If $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ and $Y_t = f(X_t)$, then $dY_t = df(X_t) = f'(X_t)(b(X_t)dt + \sigma(X_t)dW_t) + \frac{1}{2}f''(X_t)\sigma(X_t)^2dt$. Note that this last term makes this formula different from the standard chain rule.*

**Corollary 16** *If we have $Y_t = g(X_t, t)$ and $dX_t = b(X_t)dt + \sigma(X_t)dW_t$, then $dg(X_t, t) = \frac{\partial}{\partial X_t}g(X_t, t)dX_t + \frac{1}{2}\frac{\partial^2}{\partial X_t^2}g(X_t, t)\sigma(X_t)^2dt + \frac{\partial}{\partial t}g(X_t, t)dt$. (This is the same formula above except for the final time component.)*

**Corollary 17** *$E((\int_0^t f(W_s)dW_s)(\int_0^{t'} f(W_s)dW_s)) = \int_0^{\min(t,t')} E(f(W_s)^2)ds$. (This can help find covariances.)*

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:

$$d(X_tY_t) = Y_tdX_t + X_tdY_t + \sigma(X_t)\tau(Y_t)dt$$

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 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 $\widehat{X_{t+\Delta t}} \approx X_t + b(\widehat{X_t})\Delta t + \sigma(\widehat{X_t})(W_{t+\Delta t} - W_t)$. 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.

# 6 Path Integral Representations of Stochastic Differential Equations

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

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:

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

### 6.1 The Girsanov Principle

Suppose $dX_t = b(X_t)dt + dW_t$ and $dY_t = c(X_t)dt + dW_t$. Then,

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

## 7 The Fokker-Planck Equations

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

$$u(x,t) = E(f(X^x_t)) = \int_R f(y)p^x_t(y)dy$$

### 7.1 The Forward Kolmogorov Equation

The Forward Kolmogorov Equation:

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

In addition, we have the boundary conditions that $\lim_{t\to 0^+} p^x_t(y) = \delta(x-y)$, since $X_0 = x$, and $\lim_{y\to\infty} \frac{\partial}{\partial y}p^x_t(y) = 0$, 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 $\frac{\partial}{\partial t}p^x_t(y) = A^*(p^x_t(y))$.

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

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

### 7.2 The Backward Kolmogorov Equation

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

(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 $g$.)

Using this function in the formula above, we find the Backward Kolmogorov equation:

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

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