Stochastic Calculus Notes

Cliff

October 14, 2021

1 General Probability Theory

1.1 Infinite Probability Spaces

Definition 1 ($\sigma - algebra$)**.** *Let Ω be a nonempty set, and let $\mathcal{F}$ be a collection of sub-sets of Ω. We say that $\mathcal{F}$ is a $\sigma - algebra$ provided that:*

(i) the empty set ϕ belongs to F
(ii) whenever a set A belongs to F, its complement A^c also belongs to $\mathcal{F}$, and
(iii) whenever a sequence of sets $A_1, A_2, ...$ belongs to F, their union $\cup_{n=1}^{\infty} A_n$ also belongs to $\mathcal{F}$.

Definition 2 (Probability measure)**.** *Let Ω be a nonempty set, and let $\mathcal{F}$ be a collection of sub-sets of Ω. A probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathcal{F}$, assigns a number in [0,1], called the probability of A and written $\mathbb{P}(A)$. We require:*

(i) $\mathbb{P}(\Omega) = 1$, and
(ii)(countable additivity) *whenever $A_1, A_2, ...$ is a sequence of disjoint sets in $\mathcal{F}$, then*

$$\mathbb{P}\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

*The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a **probability space**.*

Definition 3 (Almost surely (a.s.))**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. If a set $A \in \mathcal{F}$ satisfies $\mathbb{P}(A) = 1$, we say that the event A occurs <u>almost surely</u>.*

1.2 Random variables and Distributions

Definition 4 (Random variable). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A random variable is a real-valued function X defined on Ω with the property that for every Borel subset B of $\mathbb{R}$, the subset of Ω given by*

$$\{X \in B\} = \{\omega \in \Omega; X(\omega) \in B\}$$

is in the $sigma - algebra \quad \mathcal{F}$. (We sometimes also permit a random variable to take the value $+\infty$ and $-\infty$)

Definition 5 (Distribution measure). *Let X be A random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution measure of X is the probability measure μ_X that assigns to each Borel subsets B of $\mathbb{R}$ The mass*

$$\mu_X(B) = \mathbb{P}\{X \in B\}$$

Theorem 1. *The relationship between cdf and distribution measure*

$$F(x) = \mathbb{P}\{X \leq x\} = \mu_X(-\infty, x], x \in \mathbb{R}$$
$$\mu_X(x, y] = \mu_X(-\infty, y] - \mu_X(-\infty, x] = F(y) - F(x)$$
$$\mu_X[a, b] = \lim_{x \to \infty} \mu_X(a - \frac{1}{n}] = F(b) - \lim_{n \to \infty} (a - \frac{1}{n})$$
$$\mu_X[a, b] = \lim_{x \to \infty} \mu_X(a - \frac{1}{n}] = \mathbb{P}\{a \leq X \leq b\} = \int_a^b f(x)dx$$

1.3 Expectations

Theorem 2. *Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

(i) If X takes only finitely many values $y_0, y_1, ..., y_n$, then

$$\int_\Omega X(\omega) d\mathbb{P}(\omega) = \sum_{k=0}^{n} y_k \mathbb{P}\{X = y_k\}$$

*(ii) **(Integrability)** The random variable X is integrable if and only if*

$$\int_\Omega |X(\omega)| d\mathbb{P}(\omega) < \infty$$

Now let Y be another random variable *(iii)* ***(Comparison)*** *If $X \leq Y$ almost surely, and if $\int_\Omega X(\omega)d\mathbb{P}(\omega)$ and $\int_\Omega Y(\omega)d\mathbb{P}(\omega)$ are defined, then*

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) \leq \int_\Omega Y(\omega)d\mathbb{P}(\omega)$$

and if $X = Y$

$$\int_\Omega X(\omega)d\mathbb{P}(\omega) = \int_\Omega Y(\omega)d\mathbb{P}(\omega)$$

(iv) ***(Linearity)*** *If α and β are real constants and X and Y are integrable, or if α and β are nonnegative constants and X and Y are nonnegative, then*

$$\int_\Omega (\alpha X(\omega) + \beta Y(\omega))d\mathbb{P}(\omega) = \alpha \int_\Omega X(\omega)d\mathbb{P}(\omega) + \beta \int_\Omega Y(\omega)d\mathbb{P}(\omega)$$

Definition 6 (Expectation)**.** *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The expectation of X is defined to be*

$$\mathbb{E} = \int_\Omega X(\omega)d\mathbb{P}(\omega)$$

This definition make sense if X is integrable or $X \geq 0$ a.s.

Theorem 3. *Let X be a random variable on probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

(i) If X takes only finitely many values $x_0, x_1, ..., x_n$, then

$$\mathbb{E}X = \sum_{k=0}^{n} x_k \mathbb{P}\{X = x_k\}$$

(ii) ***(Integrability)*** *The random variable X is integrable if and only if*

$$\mathbb{E}|X| < \infty$$

Now let Y be another random variable

(iii) ***(Comparison)*** *If $X \leq Y$ almost surely, and if $\mathbb{E}X$ and $\mathbb{E}Y$ are defined, then*

$$\mathbb{E}X \leq \mathbb{E}Y$$

and if $X = Y$

$$\mathbb{E}X = \mathbb{E}Y$$

*(iv) **(Linearity)** If α and β are real constants and X and Y are integrable, or if α and β are nonnegative constants and X and Y are nonnegative, then*

$$\mathbb{E}(\alpha X(\omega) + \beta Y(\omega)) = \alpha \mathbb{E} X(\omega) + \beta \mathbb{E} Y(\omega)$$

Definition 7 (Lebesgue measure)**.** *Let $\mathcal{B}(\mathbb{R})$ be the $\sigma-algebra$ of Borel subsets of $\mathbb{R}$. The lebesgue measure on $\mathbb{R}$, which we denote by $\mathcal{L} : \mathcal{B}(\mathbb{R}) \to [0, \infty)$, assigns to each set $B \in \mathcal{B}(\mathbb{R})$ a number in $[0, +\infty)]$ or the value ∞ so that*

(i) $\mathcal{L}[a, b] = b - a$ whenever $a \leq b$, and

(ii) if $B_1, B_2, ...$ is a sequence of disjoint sets in $\mathcal{B}(\mathbb{R})$, then we have the countable additivity property

$$\mathcal{L}\left(\cup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathcal{L}(B_n)$$

Definition 8 (Borel measurable)**.** *Let f(x) be a real-valued function defined on $\mathbb{R}$. If and only if for every Borel subsets B of $\mathbb{R}$, the set $\{x; f(x) \in B\}$ is also a Borel subset B of $\mathbb{R}$. The function $f(x)$ is called Borel-measurable.*

Theorem 4 (Comparsion of Riemann and Lebesgue intergals)**.**

(i) The Riemann intergal is defined iff the point where $f(x)$ is not continuous has Lebesgue measure equals zero.

(ii) If the Riemann intergal is defined then the Riemann and Lebesgue integral agree.

1.4 Convergence of Integrals

Definition 9 (Converge almost surely)**.**

Let $X_1, X_2, ...$ be a sequence of r.v.s defined on the same on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say $X_1, X_2, ...$ converge to another r.v. X a.s if and only if

$$\mathbb{P}(\omega \in \Omega; \lim_{n=0}^{\infty} X_n(\omega) = X(\omega)) = 1$$

or

$$\forall \varepsilon > 0, \qquad \mathbb{P}(\lim_{n\to\infty} |X_n(\omega) - X(\omega)| > \varepsilon) = 0$$

example 1 (Law of Large Numbers)**.**

$$X_1, X_2, ..., X_n \sim F_X(x)$$
$$E(X) = \mu$$

WLLN-$\bar{X} \to^{p} \mu \Leftrightarrow \lim_{n\to\infty} \mathbb{P}(|\bar{X} - \mu| > \varepsilon) = 0$

SLLN-$\bar{X} \to^{a.s.} \mu \Leftrightarrow \mathbb{P}(\lim_{n\to\infty} |\bar{X} - \mu| > \varepsilon) = 0$

Theorem 5 (Converge almost every)**.**

Let $f_1, f_2, ...$ be a sequence of real-valued, Borel-measurable functions defined on $\mathbb{R}$. Let f be another real-valued, Borel-measurable function defined on $\mathbb{R}$, we say $f_1, f_2, ... \to f$ a.e.

$$\lim_{n\to\infty} f_n = f \quad a.e.$$

if and only if

$$\forall \varepsilon > 0, \qquad \mathcal{L}(\lim_{n\to\infty} |f_n(x) - f(x)| > \varepsilon) = 0$$

example 2. *Let $f_n = \frac{n}{2\pi} e^{-\frac{nx^2}{2}}$, and it is easy to see f is a pdf of normal with $\frac{1}{n}$ variance.*

Obviously, given $f(x) = 0$

$$\mathcal{L}(\lim_{n\to\infty} f_n(x) \to 0) = 1$$

and

$$\mathcal{L}(\lim_{n\to\infty} \int_{-\infty}^{+\infty} f_n(x)dx \to 1) = 1$$

however, we know $\int_{-\infty}^{+\infty} f(x)dx = 0$, so we can conclude

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} f_n(x)dx \neq \int_{-\infty}^{+\infty} \lim_{n\to\infty} f_n(x)dx \quad (a.e.)$$

Theorem 6 (Montone convergence)**.**

Let $X_1, X_2, ...$ be a sequence of r.v.s converging almost surly to another r.v. X. If

$$0 \le X_1 \le X_2 \le ... \quad a.s.,$$

then

$$\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$$

Let $f_1, f_2, ...$ be a sequence of Borel-measurable functions converging almost surly to another Borel-measurable function f. If

$$0 \le f_1 \le f_2 \le ... \quad a.e.,$$

then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n dx = \int_{-\infty}^{\infty} f dx$$

Recall countable additivity in (i) of Thm. 3

Corollary 1. *Suppose the nonegative random r.v. X takes countable many values $x_0, x_1, ...$ Then*

$$\mathbb{E}X = \sum_{k=0}^{\infty} x_k \mathbb{P}(X = x_k)$$

Theorem 7 (Dominated convergence)**.**

Let $X_1, X_2, ...$ be a sequence of r.v.s converging almost surly to another r.v. X. If there is another r.v. Y such that $\mathbb{E}Y < \infty$ and $|X_n| \leq Y$ a.s., then

$$\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}X$$

Let $f_1, f_2, ...$ be a sequence of Borel-measurable functions converging almost surly to another Borel-measurable function f.

If there is another function $g(x)$ such that $\int_{-\infty}^{\infty} g(x) < \infty$ and $|f_n(x)| \leq g(x)$ a.e., then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n dx = \int_{-\infty}^{\infty} f dx$$

1.5 Computation of Expectations

Theorem 8.

Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let g be a Borel-measurable function on $\mathbb{R}$. Then if

$$\mathbb{E}|g(X)| = \int_{\mathbb{R}} |g(X)| d\mu_X(x) < \infty,$$

$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(X) d\mu_X(x)$$

Proof. Recall the definition of expectation $\mathbb{E} = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$ we should proof the integral w.r.t distribution measure gives the same result as integral w.r.t. probability measure on Ω.

Step 1: Indicator function

$$
\begin{aligned}
\mathbb{E}\mathbb{I}_B(X) &= \int_\Omega \mathbb{I}_B(X(\omega))d\mathbb{P}(\omega) \\
&= 0 \cdot \int_{B_c} d\mathbb{P}(\omega) + 1 \cdot \int_B d\mathbb{P}(\omega) \\
&= \mathbb{P}\{X \in B\} = \mu_X(B)
\end{aligned}
$$

in another side,

$$
\begin{aligned}
\int_\mathbb{R} \mathbb{I}_B(x)d\mu_X(x) &= 0 \cdot \int_{x \notin B} d\mu_X(x) + 1 \cdot \int_{x \in B} d\mu_X(x) \\
&= \mu_X(B)
\end{aligned}
$$

Step 2: Nonnegative simple functions

Given a simple function $g(x)$

$$
g(x) = \sum_{k=1}^n \alpha_k \mathbb{I}_{B_k}(x),
$$

$$
\mathbb{E}g(X) = \mathbb{E}\sum_{k=1}^n \alpha_k \mathbb{I}_{B_k}(x) = \sum_{k=1}^n \alpha_k \mathbb{E}\mathbb{I}_{B_k}
$$

use conclusion from step 1, it holds in simple nonnegative function.

Step 3: Nonnegative Borel-measurable function

Define Borel-sets $B = \{x; \frac{k}{2^n} \leq g(x) < \frac{k+1}{2^n}\}, k = 0, 1, 2, ..., 4^n - 1$

Then we have a partition

$$
0 < \frac{1}{2^n} < \frac{2}{2^n} < ... < \frac{4_n}{2^n} = 2^n
$$

we can see the range of partition goes to infinite.

From this construction we can get a simple function

$$
g_n(x) = \sum_{k=0}^{4^n-1} \frac{k}{2^n} \mathbb{I}_{B_{k,n}}(x)
$$

satisfy $0 \leq g_1 \leq g_2 \leq ... \leq g_n$, and from step 2, we know

$$
\mathbb{E}g_n(X) = \int_\mathbb{R} g_n(x)d\mu_X(x)
$$

Using MCT, we have

$$
\int_\mathbb{R} g(x)d\mu_X(x) = \lim_{n\to\infty} \int_\mathbb{R} g_n(x)d\mu_X(x) = \lim_{n\to\infty} \mathbb{E}g_n(x) = \mathbb{E}\lim_{n\to\infty} g_n(x) = \mathbb{E}g(X)
$$

Figure 1: Step 3

Step 4: General Borel-measurable function

Let $g(x)$ be a general Borel-measurable function,

$$g^+(x) = max\{g(x), 0\} \quad and \quad g^- = max\{-g(x), 0\}$$
$$\mathbb{E}g^+(x) = \int_{\mathbb{R}} g^+(x) d\mu_X(x) < \infty,$$
$$\mathbb{E}g^-(x) = \int_{\mathbb{R}} g^-(x) d\mu_X(x) < \infty,$$

In the end by linearity,

$$\mathbb{E}g(x) = \mathbb{E}g^+(x) - \mathbb{E}g^-(x)$$

□

Theorem 9. *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let g be a Borel-measurable function on $\mathbb{R}$. Suppose that X has a density f ($\mu_X(B) = \int_B f(x)dx$). Then if*

$$\mathbb{E}|g(X)| = \int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty$$

is finite and well-defined,

$$\mathbb{E}g(X) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

1.6 Change of measure

Theorem 10.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let Z be: (i) $\mathbb{P}\{Z > 0\} = 0$; (ii)$\mathbb{E}Z = 1$. For $A \in \mathcal{F}$, define

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega)$$

Then $\tilde{\mathbb{P}}$ is a probability measure. Furthermore, if X is a nonnegative random variable, then

$$\tilde{\mathbb{E}}X = \mathbb{E}[XZ]$$

If Z is almost surly strictly positive, we also have

$$\mathbb{E}Y = \tilde{\mathbb{E}}\left[\frac{Y}{Z}\right]$$

for every nonnegative random variable Y.

Definition 10 (Measure equivalent)**.**

Let Ω be a nonempty set and $\mathcal{F}$ a $\sigma-algebra$ of subsets of Ω. Two probability measures $\mathbb{P}$ and $\tilde{\mathbb{P}}$ on $(\Omega, \mathcal{F})$ are said to be equivalent if they agree which sets in $\mathcal{F}$ have probability zero.

Definition 11 (Radon-Nikodym derivative)**.**

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\tilde{\mathbb{P}}$ be another probability measure on $(\Omega, \mathcal{F})$ that is equivalent to $\mathbb{P}$, and let Z be an almost surly positive random variable that relates $\mathbb{P}$ and $\tilde{\mathbb{P}}$ via

$$\tilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega)$$

Then Z is called the Radon-Nikodym derivative of $\tilde{\mathbb{P}}$ w.r.t $\mathbb{P}$, and we write

$$Z = \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}$$

Theorem 11 (Radon-Nikodym)**.** *Let $\mathbb{P}$ and $\tilde{\mathbb{P}}$ be equivalent probability measures defined on $(\Omega, \mathcal{F})$. Then there exists an almost surly positive random variable Z such that $\mathbb{E}Z = 1$ and*

$$\tilde{\mathbb{P}} = \int_A Z(\omega) d\mathbb{P}(\omega) \quad \textit{for every } A \in \mathcal{F}$$

2 Information and Conditioning

2.1 Information and $\sigma - algebra$

Definition 12 (Filtration)**.**

Let Ω be a nonempty set. Let T be a fixed positive number, and assume that for each $t \in [0, T]$ there is a $\sigma - algebra \quad \mathcal{F}(t)$. Assume further that if $s \leq t$, then every set in $\mathcal{F}(s)$ is also in $\mathcal{F}(t)$. Then we call the collection of $\sigma - algebras \quad \mathcal{F}(t), 0 \leq t \leq T$, a filtration.

$$\mathcal{F}(t_1) \subset \mathcal{F}(t_2) \subset \mathcal{F}(t_3)..., \quad \textit{for } t_1 \leq t_2 \leq ...$$

Definition 13 ($\sigma(X)$)**.**

Let X be a random variable defined on a nonempty sample space Ω. The $\sigma - algebra$ generated by X, denoted $\sigma(X)$, is the collection of all subsets of Ω of the form $\{X \in B\}$, where B ranges over the Borel subsets $\mathbb{R}$, $\mathcal{B}(\mathbb{R})$.

Definition 14 ($\mathcal{G} - measurable$)**.**

Let X be a random variable defined on a nonempty sample space Ω.Let $\mathcal{G}$ be a σ-algebra of subsets of Ω.If every set in $\sigma(X)$ is also in $\mathcal{G}$, we say that X is $\mathcal{G}$ -measurable. And can be written as:

$$\sigma(X) \subset \mathcal{G}$$

Definition 15 (Adapted stochastic process)**.**

Let Ω be a nonempty sample space equipped with a filtration $\mathcal{F}(t), 0 \leq t \leq T$. Let $X(t)$ be a collection of random variables indexed by $t \in [0, T]$. We say this collection of random variables is an adapted stochastic process if, for each t, the random variable $X(t)$ is $\mathcal{F}(t) - measurable$.

2.2 Independence

Definition 16 (Independence of two σ-algebras)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G}$ and $\mathcal{H}$ be $sub - \sigma - algebras$ of $\mathcal{F}$(i.e., the sets in $\mathcal{G}$ and the sets in $\mathcal{H}$ are also in $\mathcal{F}$).We say these two σ -algebras are independent under probability measure $\mathbb{P}$ if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \textit{ for all } A \in \mathcal{G}, B \in \mathcal{H}$$

Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say these two random variables are independent if the σ-algebras they generate, $\sigma(X)$ and $\sigma(Y)$, are independent. We say that the random variable X is independent of the σ -algebra $\mathcal{G}$ if $\sigma(X)$ and $\mathcal{G}$ are independent.

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \textit{ for all } A \in \sigma(X), B \in \sigma(Y)$$

Definition 17 (General case in Independence)**.** Let$(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \ldots.$ be a sequence of sub- σ -algebras of $\mathcal{F}$. For a fixed positive integer n, we say y. that the n σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent if*

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \cdots \cdot \mathbb{P}(A_n)$$
$$\textit{for all } A_1 \in \mathcal{G}_1, A_2 \in \mathcal{G}_2, \ldots, A_n \in \mathcal{G}_n$$

Let $X_1, X_2, X_3, \ldots$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say the n random variables $X_1, X_2, \ldots, X_n$ are independent if the σ-algebras $\sigma(X_1)$ $\sigma(X_2), \ldots, \sigma(X_n)$ are independent. We say the full sequence of σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \ldots$ is independent if, for every positive integer n, the n σ-algebras $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n$ are independent. We say the full sequence of random variables $X_1, X_2, X_3, \ldots$ is independent if, for every positive integer n, the n random variables $X_1, X_2, \ldots, X_n$ are independent.

Theorem 12. *Let X and Y be independent random variables, and let f and g be Borel-measurable functions on $\mathbb{R}$. Then $f(X)$ and $g(Y)$ are independent random variables.*

For two Borel-measurable functions, their value are independent iff their independent variable are independent.

Proof. $\forall A \in \sigma(X), A = \{\omega \in \Omega; X(\omega) \in C\}$

$\forall A \in \sigma(f(X)), A = \{\omega \in \Omega; f(X(\omega)) \in E\}$

Since we have $C = \{X \in \mathbb{R}; f(X) \in E\}$, $A \in \sigma(X) \Rightarrow \sigma(f(X)) \subset \sigma(X)$.

Let B be in the σ-algebra generated by $g(Y)$. This σ -algebra is a sub σ -algebra of $\sigma(Y)$, so $B \in \sigma(Y)$. since X and Y are independent, we have $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ ☐

Definition 18 (Joint Distribution)**.** *Let X and Y be random variables. The pair of random variables (X, Y) takes values in the plane $\mathbb{R}^2$, and the joint distribution measure of (X, Y) is given by*

$$\mu_{X,Y}(C) = \mathbb{P}\{(X, Y) \in C\} \quad \textit{for all Borel sets} \quad C \subset \mathbb{R}^2$$

This is a probability measure (i.e., a way of assigning measure between 0 and 1 to subsets of $\mathbb{R}^2$ so that $\mu_{X,Y}\left(\mathbb{R}^2\right) = 1$ and the countable additivity property The joint cumulative distribution function of (X, Y) is

$$F_{X,Y}(a, b) = \mu_{X,Y}((-\infty, a] \times (-\infty, b]) = \mathbb{P}\{X \leq a, Y \leq b\}, a \in \mathbb{R}, b \in \mathbb{R}$$

We say that a nonnegative, Borel-measurable function $f_{X,Y}(x, y)$ is a joint density for the pair of random variables (X, Y) if

$$\mu_{X,Y}(C) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}_C(x, y) f_{X,Y}(x, y) dy dx$$

for all Borel sets $C \subset \mathbb{R}^2$

Definition 19 (Joint density function)**.** *We say that a nonnegative, Borel-measurable functioon $f_{X,Y}(x, y)$ is a joint density for the pair of random variables (X,Y) if*

$$\mu_{X,Y}(C) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{I}_C(x, y) f_{X,Y}(x, y) dy dx \textit{ for all Borel sets } C \subset \mathbb{R}^2$$

Condition holds iff

$$F_{X,Y}(a, b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{X,Y}(x, y) dy dx \textit{ for all } a \in \mathbb{R}, b \in \mathbb{R}$$

Theorem 13. *Let X and Y be random variables. The following conditions are equivalent.*

(i) X and Y are independent.

(ii) The joint distribution measure factors:

$$\mu_{X,Y}(A \times B) = \mu_X(A) \cdot \mu_Y(B) \text{ for all Borel subsets } A \subset \mathbb{R}, B \subset \mathbb{R}$$

(iii) The joint cumulative distribution factors:

$$F_{X,Y}(a,b) = F_X(a) \cdot F_Y(b) \text{ for all } a \in \mathbb{R}, b \in \mathbb{R}$$

(iv) The joint moment-generating function factors:

$$\mathbb{E}e^{uX+vY} = \mathbb{E}e^{uX} \cdot \mathbb{E}e^{vY}$$

for all $u \in \mathbb{R}, v \in \mathbb{R}$ for which the expectations are finite.

If there is a joint density, each of the conditions above is equivalent to the following.

(iv) The joint density factors:

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \text{ for almost every } x \in \mathbb{R}, y \in \mathbb{R}$$

The conditions above imply but are not equivalent to the following.
(vi) The expectation factors:

$$\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y, \\ \text{provided } \mathbb{E}|XY| < \infty$$

2.3 Discrete-Time Martingale

Definition 20 (Martingale)**.** *$\{X_n, n \geq 0\}$ is a martingale (or sub- or super-martingale) w.r.t. $\{\mathcal{F}_n\}$ if*

1. *$E|X_n| < \infty$*
2. *$\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$*
3. *$E(X_{n+1} \mid \mathcal{F}_n) = X_n$ (or $\geq X_n$ or $\leq X_n$) for all n*

example 3 (Doob's martingale)**.** *Let Z be a r.v with $E(|Z|) < \infty$, and $Y_n = E(Z|\mathcal{F}_n)$, we have Y_n is a martingale.*

Proof.

1.

$$E(|Y_n|) = E(|E(Z|\mathcal{F}_{n-1})|) \leq E(E(|Z||\mathcal{F}_{n-1})) = E(||Z|) < \infty \tag{1}$$

2.

$$Y_n \in \mathcal{F}_n \tag{2}$$

3.

$$E(Y_{n+1}|\mathcal{F}_n) = E(E(Z|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(Z|\mathcal{F}_n) = Y_n \tag{3}$$

□

Theorem 14 (Doob's decomposition theorem)**.** $\{Y_n\}$ *is a submartingale, then*

$$Y_n = Y_0 + M_n + A_n, \quad n \geq 0$$

$$Y_n = Y_0 + \sum_{k=1}^{n}(Y_k - Y_{k-1})$$

$$Y_0 + \sum_{k=1}^{n}\left(Y_k - E\left(Y_k \mid \mathcal{F}_{k-1}\right)\right) + \sum_{k=1}^{n}\left(E\left(Y_k \mid \mathcal{F}_{k-1}\right) - Y_{k-1}\right)$$

where
- M_n *is a martingale with* $M_0 = 0$
- A_n *is an increasing predictable process with* $A_0 = 0$
- *This decomposition is unique.*

3 Brownian Motion

3.1 Reflection Principle

Theorem 15 (Reflecion Equality)**.**

$$P(W(t) < \omega, \tau_m < t) = P(W(t) > 2m - \omega)$$

Reflection Equality.

$$\begin{aligned} P(W(t) > 2m - \omega) =& P(W(t) > 2m - \omega | \tau_m < t)P(\tau_m < t) \\ &+ P(W(t) > 2m - \omega | \tau_m \geq t)P(\tau_m \geq t) \end{aligned}$$

where $P(W(t) > 2m - \omega | \tau_m \geq t)P(\tau_m \geq t) = 0$
So,

$$\begin{aligned} P(W(t) > 2m - \omega) =& P(W(t) > 2m - \omega | \tau_m < t)P(\tau_m < t) \\ &= P(W(t) < \omega | \tau_m < t)P(\tau_m < t)[Reflection] \\ &= P(W(t) < \omega, \tau_m < t) \end{aligned}$$

□

3.2 Stochastic Calculus

3.2.1 *Itô* Integrals

Define an function $f : f(x, y)$ and its difference $df(x, y)$,

$$\begin{aligned} df =& f_x dx + f_y dy \\ &+ \frac{1}{2} f_{xx} dxdx + f_{xy} dxdy + \frac{1}{2} f_{yy} dydy \\ &+ \frac{1}{3!} f_{xxx} (dx)^3 + \frac{2}{3!} f_{xxy} dxdxdy + \frac{2}{3!} f_{xyy} dxdydy + \frac{1}{3!} f_{yyy} (dy)^3 + \ldots \end{aligned} \tag{4}$$

if x, y are continuous differentiate functions, $(dx)^2, dxdy, (dy)^2 = 0$,

if x is an $It\hat{o}$ process and y is a continuous differentiate function, $dxdy, (dy)^2 = 0$ and $(dx)^2 \neq 0$

if x, y are $It\hat{o}$ processes, $(dx)^2, dxdy, (dy)^2 \neq 0$

And $(dx)^3, (dy)^3, (dxdxdy), (dxdydy) = 0$ in all three above cases.

3.2.2 BSM partial differential equation

Considering a hedging portfolio $X(t)$ with stock and money account, and an option $c(t, S_t)$.

Assume the underlying stock is S_t and its dynamics defined as fellows,

$$dS_t = \alpha S_t dt + \sigma S_t dW_t \tag{5}$$

then let's denote the stocks holding at t,

$$dX_t = \underbrace{\Delta_t dS_t}_{\text{Earnings from the stock price}} + \underbrace{r(X_t - \Delta_t S_t)dt}_{\text{Earnings from the money account}} \tag{6}$$

We can get and pde that

$$d(e^{-rt}X_t) = d(e^{-rt}c(t, S_t)) \tag{7}$$

and combing the **initial condition** $X_0 = c(0, S_0)$,

$$X_T = X_0 + \int_0^T d(e^{-rt}X_t) = c(0, S_0) + \int_0^T d(e^{-rt}c(t, S_t)) = c(t, S_t) \tag{8}$$

Finally, we can compute $d(e^{-rt}X_t) = d(e^{-rt}c(t, S_t))$

$$d(e^{-rt}X_t) = \Delta_t \left[(\alpha - r)e^{-rt}S_t dt + \sigma e^{-rt}S_t dW_t\right]$$

$$\begin{aligned}
&d(e^{-rt}X_t) = \Delta_t \left[(\alpha - r)e^{-rt}S_t dt + \sigma e^{-rt}S_t dW_t\right] \\
&= \Delta_t d(e^{-rt}S_t) = d(e^{-rt}c(t, S_t)) \\
&= e^{-rt}\left[-rc(t, S_t) + c_t(t, S_t) + \alpha S_t c_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 c_{xx}(t, S_t)\right] dt \\
&+ e^{-rt}\sigma S_t c_x(t, S_t) dW_t
\end{aligned} \tag{9}$$

we can simplify the equation as

$$
\begin{aligned}
&\Delta_t \left[(\alpha - r)e^{-rt} S_t dt + \sigma e^{-rt} S_t dW_t\right] \\
&= e^{-rt} \left[-rc(t, S_t) + c_t(t, S_t) + \alpha S_t c_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 c_{xx}(t, S_t)\right] dt \\
&+ e^{-rt} \sigma S_t c_x(t, S_t) dW_t
\end{aligned} \tag{10}
$$

The above equation means we can hedge the underlying randomness through holding $c_x(t, S_t)$ shares of underlying, which is $\Delta_t = c_x(t, S_t)$

after substitute the $c_x(t, S_t)$ into Δ_t and cancel dW_t terms,

$$
\begin{aligned}
&c_x(t, S_t) \left[(\alpha - r)e^{-rt} S_t dt\right] \\
&= e^{-rt} \left[-rc(t, S_t) + c_t(t, S_t) + \alpha S_t c_x(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 c_{xx}(t, S_t)\right] dt
\end{aligned} \tag{11}
$$

becomes

$$
\begin{aligned}
&- rS_t c_x(t, S_t) dt \\
&= \left[-rc(t, S_t) + c_t(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 c_{xx}(t, S_t)\right] dt
\end{aligned} \tag{12}
$$

we reformat it and get the *Black-Scholes-Merton Equation*.

$$
-rc(t, S_t) = rS_t c_x(t, S_t) + c_t(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 c_{xx}(t, S_t) \tag{13}
$$

with **terminal condition** $c(T, x) = (x - K)^+$.