AN INTRODUCTION TO THE STOCHASTIC INTEGRAL

MATT OLSON

ABSTRACT. This paper gives an elementary introduction to the development of the stochastic integral. I aim to provide some of the foundations for someone who wants to begin the study of stochastic calculus, which is of great importance in the theory of options pricing.

1. Introduction

Stochastic calculus is now one of the central tools in modern Mathematical Finance. Its beginnings can be traced back to L. Bachelier's 1990 dissertation *Thorie de la spculation* in which he modeled stock prices with Brownian motion. Nearly one hundred years later Robert Miller and Myron Scholes won the Nobel Prize using stochastic calculus and arbitrage pricing to derive the famed Black-Scholes equation. In this paper I will provide a hopefully gentle introduction to stochastic calculus via the development of the stochastic integral.

I have found that in the literature there is a great divide between those introductory texts which are only accessible to PhD's on the one hand, and those which lack rigor altogether and are directed towards traders. Armed with some basic analysis and probability this presentation should be accessible at the undergraduate level.

2. Preliminaries

Definition 2.1. A **Brownian motion** on $[0,1]$ is a stochastic process $\{B_t : 0 \leq t \leq 1\}$ on some probability space $(\Omega, \mathscr{F}, P)$ adapted to $\{\mathscr{F}_t\}$ with the following properties:

(i) $B_0 = 0$
(ii) The random variable $B_t - B_s$ is independent of $\mathscr{F}_s$ for any $s \leq t$.
(iii) For any $0 \leq s \leq t \leq 1$ $B_t - B_s$ is normally distributed with mean 0 and variance $t - s$.
(iv) With probability 1 $B_t(\omega)$ is a continuous function of t.

A filtration $\{\mathscr{F}_t\}_{t\geq 0}$ is a family of sub-sigma algebras of some sigma-algebra $\mathscr{F}$ with the property that if $s < t$ then $\mathscr{F}_s \subset \mathscr{F}_t$. Saying a process $\{X_t : 0 \leq t < \infty\}$ is adapted to $\{\mathscr{F}_t\}_{t\geq 0}$ means that for any t X_t is $\mathscr{F}_t$ measurable.

Definition 2.2. The process $\{X_t : 0 \leq t < \infty\}$ adapted to the filtration $\{\mathscr{F}_t\}_{t\geq 0}$ is a **martingale** if the following conditions hold:

(i) $E[|X_t|] < \infty$ for all $0 \leq t < \infty$
(ii) $E[X_t|\mathscr{F}_s] = X_s$ for all $0 \leq s \leq t < \infty$

Remark 2.3. An equivalent assertion to Property (iii) is that for any $0 \le t_1 \le \cdots \le t_n \le 1$ the random variables

$$B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$$

are independent.

Proposition 2.4. *A Brownian motion $\{B_t\}$ is a martingale with respect to its filtration*

Proof. Let $0 \le s \le t \le 1$. We can decompose the conditional expectation into

$$E[B_t|\mathscr{F}_s] = E[B_t - B_s|\mathscr{F}_s] + E[B_s|\mathscr{F}_s].$$

By Property (ii) of Brownian motion $B_t - B_s$ is independent of $\mathscr{F}_s$ so $E[B_t - B_s|\mathscr{F}_s] = E[B_t - B_s] = 0$. Also, B_s is $\mathscr{F}_s$ measurable so $E[B_s|\mathscr{F}_s] = B_s$. Combining these results we have $E[B_t|\mathscr{F}_s] = B_s$. □

Later on we will need to use the following result which we will prove in Proposition 2.5:

$$\lim_{\|P\| \to 0} \sum_{i=1}^{n} (B_{t_i} - B_{t_{i-1}})^2 = T$$

where $P = \{t_0, \ldots, t_n\}$ is any partition of [0,T] and the convergence is in $L^2(\Omega)$. This limit is called the **quadratic variation** of the Brownian motion and is one measure of its volatility. Most functions we see in ordinary calculus have zero quadratic variation, and in fact it is not hard to see that any function with a continuous derivative has zero quadratic variation.

Proposition 2.5. *The quadratic variation of B_t over $[0, T]$ is T.*

Proof. Consider a partition $P = \{t_0, \ldots, t_n\}$. To simplify notation define a new variable $X_i = B_{t_i} - B_{t_{i-1}}$. Recall that by Property (ii) of Brownian motion $X_i = B_{t_i} - B_{t_{i-1}}$ follows a normal distribution with mean 0 and variance $(t_i - t_{i-1})$, so that $E[X_i^2] = (t_i - t_{i-1})$. Therefore,

$$E[\sum_{i=1}^{n} X_i^2] = \sum_{i=1}^{n} E[X_i^2] = \sum_{i=1}^{n} (t_i - t_{i-1}) = T. \tag{2.6}$$

With a little work one can show that $E[X_i^4] = 2(t_i - t_{i-1})^2$, which in turn implies that $Var[X_i^2] = E[(X_i^2)^2] - (E[X_i^2])^2 = (t_i - t_{i-1})^2$. Now, using the independence of the X_i,

$$\begin{aligned} E[(\sum_{i=1}^{n} X_i^2 - T)^2] = Var[\sum_{i=1}^{n} X_i^2] = \sum_{i=1}^{n} Var[X_i^2] &= \sum_{i=1}^{n} (t_i - t_{i-1})^2 \\ &\le T\|P\| \end{aligned} \tag{2.7}$$

If we choose partitions such that $\|P\| \to 0$ then $E[(\sum_{i=1}^{n} X_i^2 - T)^2]$ converges to 0. □

3. Stochastic Integral for Simple Functions

3.1. **Motivation.** Imagine we model the price of an asset as a Brownian motion with value B_t at time t [1]. Suppose we are allowed to trade our asset only at the following times: $0 = t_0 < t_1 < \cdots < t_n = 1$. At time t_k we can choose to hold X_k shares of our asset, and we must hold these shares up until the next time period t_{k+1}. Note that in making the decision to hold X_k shares we are only allowed to use information up to that time - we cannot predict future price movements. The change in the value of our portfolio between time t_{k-1} and time t_k is $X_{k-1}(B_{t_k} - B_{t_{k-1}})$, which is simply the change in price multiplied by the number of shares we owned. For instance, if at the beginning of the period we are holding 5 shares for $5 a piece but by the end of the period the shares are worth $7 a piece, over that interval we have made $10. It is clear then that the change in our wealth over the time period $[0, 1]$ is given by:

$$\sum_{i=1}^{n} X_{i-1}(B_{t_i} - B_{t_{i-1}})$$

It is our ultimate goal to consider this quantity as we allow for trading in continuous time, i.e. we can buy or sell an asset at any $t \in [0, 1]$.

3.2. **The Integral.**

Definition 3.1. A **simple process** $f(t,\omega)$ is a stochastic process of the form $f(t,\omega) = \sum_{i=1}^{n} \xi_{i-1}(\omega) 1_{[t_{i-1},t_i)}(t)$ where $0 = t_0 < t_1 < \cdots < t_n = 1$ is a partition of [0,1] and $\xi_{i-1}(\omega)$ is a $\mathscr{F}_{t_{i-1}}$ measurable random variable. [2]

Intuitively, a simple process can be thought of as a step function on [0,1] with each step taking on a random value. A simple process is analogous to the trading situation above where we chose certain amounts of stocks to hold over a time interval, where one's decision was based solely on past knowledge.

Notation 3.2. On the partition $0 = t_0 < t_1 < \cdots < t_n = 1$ let $S(t_0, \ldots, t_n)$ denote the class of simple stochastic processes adapted to $\{\mathscr{F}_t\}$ with the additional requirement that if $f(t,\omega) \in S(t_0, \ldots, t_n)$ then $\|f\|^2_{[0,1]\times\Omega} < \infty$. [3] Finally, we let $S = \cup S(t_0, \ldots, t_n)$, the union taken over all partitions of [0,1].

Definition 3.3. The **stochastic integral** for $f(t,\omega) \in S$, more specifically $f(t,\omega) \in S(t_0, \ldots, t_n)$, from 0 to T is $\int_0^T f dB = \sum_{i=1}^{k-1} \xi_{i-1}(B_{t_i} - B_{t_{i-1}}) + \xi_k(B_T - B_{t_k})$ for $T \in [t_k, t_{k+1}]$. We also use the notation $\int_0^T f dB = I(T,\omega,f)$.

From the definition it is not hard to see that $I : S \to L^2(\Omega)$ is linear, i.e. $I(t,\omega, af + bg) = aI(t,\omega,f) + bI(t,\omega,g)$ for $a, b \in \mathbb{R}$. The following proposition is of crucial importance to the extension of the integral to a wider class of functions.

Proposition 3.4. *For $f(t,\omega) \in S$ we have $\|I(t,\omega,f)\|^2_\Omega = \|f\|^2_{[0,t]\times\Omega}$. In other words, $I : S \to L^2(\Omega)$ is an isometric mapping.*

---

[1] Strictly speaking a Brownian motion is not an appropriate model for an asset price such as a stock which takes only non-negative values. Generally one considers a geometric Brownian motion - however our example does provide the necessary motivation for better developed models

[2] From now on we will suppress the ω argument in $\xi_{i-1}(\omega)$.

[3] We take $\|f\|^2_{[0,1]\times\Omega} = \int_0^1 E[f^2(t,\omega)]dt$, where $E[f(t,\omega)]$ is the expectation.

Proof. In order to make life easier, but with no loss of generality, we take $t = t_k$ so that t coincides with some partition point. Define $X_i = B_{t_i} - B_{t_{i-1}}$. Now $I^2(t,\omega,f) = \sum_{i=1}^{k} \xi_{i-1}^2 X_i^2 + \sum_{i\neq j} \xi_{i-1}\xi_{j-1}X_iX_j$. One can argue that from the integrability condition on $f(t,\omega)$ and the Cauchy-Schwartz inequality the expectations of both these sums exist. To that end we consider $E[\xi_{i-1}\xi_{j-1}X_iX_j]$ for $i < j$. Since for $i < j$ $\xi_{i-1}\xi_{j-1}X_i$ is $\mathscr{F}_{j-1}$ measurable, $E[\xi_{i-1}\xi_{j-1}X_iX_j] = E[E[\xi_{i-1}\xi_{j-1}X_iX_j|\mathscr{F}_{j-1}] = E[\xi_{i-1}\xi_{j-1}X_iE[X_j|\mathscr{F}_{j-1}]]$. But because by Property (iii) of Brownian motion, X_j is independent of $\mathscr{F}_{j-1}$, so $E[X_j|\mathscr{F}_{j-1}] = E[X_j] = 0$ and $E[\xi_{i-1}\xi_{j-1}X_iX_j] = 0$. Thus we have:

$$E[\sum_{i\neq j} \xi_{i-1}\xi_{j-1}X_iX_j] = 0 \tag{3.5}$$

Furthermore, $E[\xi_{i-1}^2 X_i^2] = E[\xi_{i-1}^2]E[X_i^2] = E[\xi_{i-1}^2](t_i - t_{i-1})$ and

$$\sum_{i=1}^{k} E[\xi_{i-1}^2 X_i^2] = \sum_{i=1}^{k} E[\xi_{i-1}^2](t_i - t_{i-1}) = \|f\|^2_{[0,t]\times\Omega} \tag{3.6}$$

so that combining 3.5 and 3.6 $\|I(t,\omega,f)\|^2_\Omega \equiv E[I^2(t,\omega,f)] = \sum_{i=1}^{k} E[\xi_{i-1}^2 X_i^2] = \|f\|^2_{[0,t]\times\Omega}$ □

Proposition 3.7. *If $f(t,\omega) \in S$, then $I(t,\omega,f)$ is a martingale with respect to the $\{\mathscr{F}_t\}$ filtration.*

Proof. Since $f(t,\omega) \in S$ there is a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ such that $f(t,\omega) = \sum_{i=1}^{n} \xi_{i-1}(\omega)1_{[t_{i-1},t_i)}(t)$. Take $0 \le s \le t \le 1$. Either s and t are in the same partition interval, or they are in different intervals - we consider the more difficult case when they are in different intervals. Assume $t_k \le s \le t_{k+1}$ and $t_l \le t \le t_{l+1}$ with $k < l$. Simply by the linearity of conditional expectation

$$\begin{aligned} E[I(t,\omega,f)|\mathscr{F}_s] &= E[\sum_{i=1}^{k} \xi_{i-1}(B_{t_i} - B_{t_{i-1}})|\mathscr{F}_s] + E[\xi_k(B_{t_{k+1}} - B_{t_k})|\mathscr{F}_s] \\ &+E[\sum_{i=k+2}^{l} \xi_{i-1}(B_{t_i} - B_{t_{i-1}})|\mathscr{F}_s] + E[\xi_l(B_t - B_{t_l})|\mathscr{F}_s]. \end{aligned} \tag{3.8}$$

Consider the second term on the RHS of (3.8):

$$\begin{aligned} E[\xi_k(B_{t_{k+1}} - B_{t_k})|\mathscr{F}_s] &= E[\xi_k B_{t_{k+1}}|\mathscr{F}_s] - E[\xi_k B_{t_k}|\mathscr{F}_s] \\ &= \xi_k E[B_{t_{k+1}}|\mathscr{F}_s] - \xi_k E[B_{t_k}|\mathscr{F}_s] \\ &= \xi_k(B_{t_s} - B_{t_k}) \end{aligned} \tag{3.9}$$

The last step follows from the martingale property of Brownian motion. Now we look at the fourth term on the RHS of (3.8), keeping in mind the Tower Property of conditional expectation:

$$E[\xi_l(B_t - B_{t_l})|\mathscr{F}_s] = E[E[\xi_l(B_t - B_{t_l})|\mathscr{F}_l]|\mathscr{F}_s]$$
$$= E[\xi_l E[(B_t - B_{t_l})]|\mathscr{F}_s]$$
$$= 0 \tag{3.10}$$

By similar types of arguments we can show

$$E[\sum_{i=1}^{k} \xi_{i-1}(B_{t_i} - B_{t_{i-1}})|\mathscr{F}_s] = \sum_{i=1}^{k} \xi_{i-1}(B_{t_i} - B_{t_{i-1}}) \tag{3.11}$$

$$E[\sum_{i=k+2}^{l} \xi_{i-1}(B_{t_i} - B_{t_{i-1}})|\mathscr{F}_s] = 0 \tag{3.12}$$

Together (3.9)-(3.12) imply $E[I(t,\omega,f)|\mathscr{F}_s] = I(s,\omega,f)$. □

Example 3.13. A step function (in the usual sense from analysis) is a special type of simple stochastic process - we can think of the steps as "random" but taking certain constant values with probability one. For instance, let $f(t) = \sum_{i=1}^{n} \xi_{i-1} 1_{[t_{i-1},t_i)}(t)$ where in this case each ξ_{i-1} can be thought of as a true constant. By definition $\int_0^1 f dB = \sum_{i=1}^{n} \xi_{i-1}(B_{t_i} - B_{t_{i-1}})$, which is a linear combinations of independent normal random variables. Since a linear combination of independent normal random variables is again a normally distributed random variable, it remains to compute the mean and variance to completely characterize the distribution of $I(1,\omega,f)$.

By Proposition 3.7 $I(1,\omega,f)$ is a martingale, and consequently for any $s,t \in [0,1]$, $E[I(s,\omega,f)] = E[I(t,\omega,f)]$. When $s = 0$, $E[I(0,\omega,f)] = 0$, so clearly $E[I(t,\omega,f)] = 0$ for arbitrary t. The formula for variance simplifies since the mean is zero:

$$Var[I(t,\omega,f)] = E[I^2(t,\omega,f)] = \int_0^1 E[f^2(t,\omega)]dt = \sum_{i=1}^{k} E[\xi_{i-1}^2](t_i - t_{i-1}) = \sum_{i=1}^{k} \xi_{i-1}^2(t_i - t_{i-1})$$

by the fact that ξ_{i-1} is non-random. In summary, the distribution of $\int_0^1 f dB$ is $N(0, \sum_{i=1}^{k} \xi_{i-1}^2(t_i - t_{i-1}))$.

4. An extension of the Stochastic Integral

In this section we construct the extension of the stochastic integral from the class S of simple adapted stochastic processes to the closure of S, $\bar{S}$, where the closure is with respect to the usual norm of $L^2([0,1]\times\Omega)$. It can be shown that $\bar{S}$ contains the class of square integrable adapted functions, which is of particular interest [2].

To begin, we note that any function $f(t,\omega) \in \bar{S}$ can be approximated by a sequence of functions $\{f_n(t,\omega)\} \in S$, i.e. $\lim_{n\to\infty} \|f(t,\omega) - f_n(t,\omega)\|_{[0,1]\times\Omega} = 0$. Note that this convergence tells us that $f_n(t,\omega)$ is a Cauchy sequence, but then $\{I(t,\omega,f_n)\}$ must also be a Cauchy sequence since by Proposition 3.4

$$\|I(t,\omega,f_m) - I(t,\omega,f_n)\|_\Omega = \|f_m - f_n\|_{[0,1]\times\Omega}.$$

By the completeness of $L^2(\Omega)$ the following definition makes sense:

Definition 4.1. $I(t,\omega,f) = \lim_{n\to\infty} I(t,\omega,f_n)$ is the **stochastic integral** of $f(t,\omega) \in \bar{S}$, where $\{f_n(t,\omega)\}$ is the sequence in S mentioned in the prior discusison.

Proposition 4.2. *(Ito Isometry) The mapping $I : \bar{S} \to L^2(\Omega)$ is an isometry.*

Proof. From triangle inequality and Proposition 3.4, we have

$$\| \|I(t,\omega,f)\|_\Omega - \|f_n(t,\omega)\|_{[0,1]\times\Omega}\| \leq \|I(t,\omega,f) - I(t,\omega,f_n)\| \to 0$$

so that $\|I(t,\omega,f)\|_\Omega = \lim \|f_n(t,\omega)\|_{[0,t]\times\Omega} = \|f(t,\omega)\|_{[0,t]\times\Omega}$ □

A similar type of argument $I(t,\omega,f)$ is linear in both $f(t,\omega)$ and t. We state as a fact that $I(t,\omega,f)$ is continuous in t and adapted to $\{\mathscr{F}_t\}$.

Remark 4.3. Recall that any isometric mapping $A : X \to Y$ between two real normed inner product spaces preserves inner products: for any $x,z \in X$, $\langle Ax, Ay\rangle = \langle x,y\rangle$. Therefore $E[I(t,\omega,f)I(t,\omega,g)] = \int_0^t E[fg]dt$

Proposition 4.4. *$I(t,\omega,f)$ is a martingale in the t argument.*

Proof. We follow the proof of this found in [2]. Let $f \in \bar{S}$, $0 \leq s \leq t \leq 1$, and define $X_t = \int_0^t f dB$. By the linearity of $\int_0^t f dB$ and conditional expectation we can write

$$E[X_t|\mathscr{F}_s] = E[X_s|\mathscr{F}_s] + E[X_t - X_s|\mathscr{F}_s] \tag{4.5}$$

Since X_s is $\mathscr{F}_s$ measurable, $E[X_s|\mathscr{F}_s] = X_s$ so from (4.5) it suffices to show that $E[X_t - X_s|\mathscr{F}_s] = 0$. Now pick a sequence of $f_n \in S$ that converge in $L^2([0,1]\times\Omega)$ to $f \in \bar{S}$ and call $X_t^{(n)} = \int_0^t f_n dB$. Now

$$X_t - X_s = (X_t - X_t^{(n)}) + (X_t^{(n)} - X_s^{(n)}) + (X_s^{(n)} - X_s) \tag{4.6}$$

From the previous section we know $X_t^{(n)}$ is a martingale, $E[X_t^{(n)} - X_s^{(n)}|\mathscr{F}_s] = 0$. Now consider conditional expectation of the first term in (4.6):

$$E[X_t - X_t^{(n)}|\mathscr{F}_s] = E[\int_0^t (f - f_n)dB|\mathscr{F}_s].$$

Also,

$$\begin{aligned} E[(E[\int_0^t (f-f_n)dB|\mathscr{F}_s])^2] \leq& E[E[(\int_0^t (f-f_n)dB)^2|\mathscr{F}_s]] = E[(\int_0^t (f-f_n)dB)^2] \\ =& \|f_n - f\|^2_{[0,t]\times\Omega} \leq \|f_n - f\|^2_{[0,1]\times\Omega} \to 0 \end{aligned}$$

where the first inequality follows from Jensen's inequality, the first equality is a result of the law of iterated expectation, and the second equality follows from the Ito isometry. From this we can conclude $E[\int_0^t (f - f_n)dB|\mathscr{F}_s] = E[X_t^{(n)} - X_t|\mathscr{F}_s]$ converges to zero with probability one, and the same reasoning applies to $E[X_s^{(n)} - X_s|\mathscr{F}_s]$. Combining these results we see $E[X_t - X_s|\mathscr{F}_s] = 0$ as desired. □

Example 4.7. We assumed at the beginning the B_t was adapted to $\{\mathscr{F}_t\}$ and by a straightforward computation $\|B_t\|^2_{[0,1]\times\Omega} = \int_0^1 E[B_t^2]dt = \int_0^1 tdt < \infty$ so that we can write $\int_0^T B_t\, dB$ for $0 \le T \le 1$. We will explicitly calculate this stochastic integral from the definition.

To start off with, we will find an approximating sequence of functions $B_t^n \in S$ of B_t on $[0, T]$. Pulling a rabbit out of a hat, we find one such convenient sequence of functions is

$$\sum_{i=1}^{n} B_{\frac{(i-1)T}{n}} 1[\frac{(i-1)T}{n}, \frac{iT}{n})$$

In words, take a partition of $[0, 1]$ with points $0 < \frac{T}{n} < \frac{2T}{n} < \cdots < 1$ and on each interval $[\frac{(i-1)T}{n}, \frac{iT}{n})$ let $B_t^{(n)}$ take the value that B_t takes at the first time in that interval. One can verify that $B_t^{(n)} \in S$ for every n, and moreover $\|B_t^{(n)} - B_t\|_{[0,1]\times\Omega} \to 0$. By definition of the stochastic integral

$$\int_0^T B_t dB = \lim_{n\to\infty} \int_0^T B_t^{(n)} dB = \lim_{n\to\infty} \sum_{i=1}^{n} B_{\frac{(i-1)T}{n}} (B_{\frac{iT}{n}} - B_{\frac{(i-1)T}{n}}). \tag{4.8}$$

We focus on this last sum. By expanding the sum and rearranging we can show

$$\sum_{i=1}^{n} B_{\frac{(i-1)T}{n}} (B_{\frac{iT}{n}} - B_{\frac{(i-1)T}{n}}) = \frac{1}{2}B_T^2 - \frac{1}{2}\sum_{i=1}^{n} (B_{\frac{iT}{n}} - B_{\frac{(i-1)T}{n}}) \tag{4.9}$$

But from a previous result we know that $\lim_{n\to\infty} \sum_{i=1}^{n} (B_{\frac{(iT)}{n}} - B_{\frac{(i-1)T}{n}})^2 = T$. Thus, combining (4.8) and (4.9)

$$\int_0^T B_t dB = \frac{1}{2}B_T^2 - \frac{1}{2}T.$$

Note the extra $\frac{1}{2}T$ term which differs from the usual Reimann-Steiltjes integral: $\int gdg = \frac{1}{2}g^2$. As an exercise confirm that $\frac{1}{2}B_T^2 - \frac{1}{2}T$ is a martingale as we would expect.

Appendix A. Conditional Expectation

Definition A.1. For an integrable random variable X on $(\Omega, \mathscr{F}, P)$ and we define the **conditional expectation** of X with respect to $\mathscr{G} \subset \mathscr{F}$ to be the random variable $E[X|\mathscr{G}]$ with the following properties:

(i) $E[X|\mathscr{G}]$ is $\mathscr{G}$ measurable
(ii) For any $A \in \mathscr{G}$ $\int_A XdP = \int_A E[X|\mathscr{G}]dP$

The existence of such a random variable can be shown using the *Radon-Nikodym* theorem. Intuitively, the conditional expectation represents our best guess about the expectation of random variable given some information. The following is a list of facts about conditional expectation that we will use numerous times.

Properties of Conditional Expectation

1. **Law of Iterated Expectation** : $E[E[X|\mathscr{G}]] = E[X]$
2. If $E|XY| < \infty$ and X is $\mathscr{G}$ measurable, then $E[XY|\mathscr{G}] = XE[Y|\mathscr{G}]$
3. If X is independent of $\mathscr{G}$, then $E[X|\mathscr{G}] = E[X]$
4. **Tower Property**: If $\mathscr{G}_2 \subset \mathscr{G}_1 \subset \mathscr{F}$, then $E[E[X|\mathscr{G}_1]|\mathscr{G}_2]] = E[X|\mathscr{G}_2]$

(5) **Conditional Jensen Inequality**: If $\phi(x) : \mathbb{R} \to \mathbb{R}$ is convex and $E[\phi(X)] < \infty$, then $\phi(E[X|\mathscr{G}]) \leq E[\phi(X)|\mathscr{G}]$

References

[1] Kiyosi Ito. Stochastic Integral. Proc. Imp. Acad. Volume 20, Number 8 (1944), 519-524.
[2] Hui-Hsiung Kuo. Introduction to Stochastic Integration. Springer, 2006.
[3] Steve Shreve. Stochastic Calculus for Finance II. Springer, 2004.
[4] J. Steele. Stochastic Calculus and Financial Applications Springer, 2001.