1 Ito's formula part II

Here we generalize Ito's formula to allow for real-valued functions of two variables, where time $t \geq 0$ is one of the two; $f = f(t,x)$; $f : \mathbf{R}_+ \times \mathbf{R} \rightarrow \mathbf{R}$. We also introduce Ito processes, and generalize Ito's formula to cover them too. Finally, we intoduce more formally, stochastic differential equations and Ito diffusions.

1.1 Ito's formula for $f = f(t,x)$

If $X(t) = \sigma B(t) + \mu t,\ t \geq 0$ is BM with drift and variance term, then it is clear that $dX(t) = \sigma dB(t) + \mu dt$. But such an obvious result can not be derived directly via Ito's formula for one-variable functions $f = f(x)$: There is no such function that maps $\{B(t)\}$ to $\{X(t)\}$. What is needed is a function of two variables $f = f(t,x)$, in which the time variable t is included. For example, here we need $f(t,x) = \sigma x + \mu t$. It turns out that Ito's formula can be extended to cover such functions, and the proof (left out) is a simple modification of the proof for $f = f(x)$ using the more general Taylor's series expansion

$$f(t+s, x+h) = f(t,x) + h\frac{\partial f}{\partial x}(t,x) + s\frac{\partial f}{\partial t}(t,x) + \frac{1}{2}h^2\frac{\partial^2 f}{\partial x^2}(t,x) + ERROR$$

Theorem 1.1 *If $f = f(t,x)$ is a function such that both $\frac{\partial f}{\partial t}(t,x)$ and $\frac{\partial^2 f}{\partial x^2}(t,x)$ are continuous, then*

$$f(t,B(t)) = f(0,0) + \int_0^t \frac{\partial f}{\partial x}(s,B(s))dB(s) + \int_0^t \frac{\partial f}{\partial t}(s,B(s))ds + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s,B(s))ds.$$

In differential form the formula is given by

$$df(t,B(t)) = \frac{\partial f}{\partial x}dB(t) + (\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2})dt.$$

Examples

1. *BM with drift and variance term:* As our first example, take $f(t,x) = \sigma x + \mu t$, and let $X(t) = \sigma B(t) + \mu t$. Then $\frac{\partial f}{\partial x} = \sigma$, $\frac{\partial f}{\partial t} = \mu$ and $\frac{\partial^2 f}{\partial x^2} = 0$ yielding (in differential form) $dX(t) = \sigma dB(t) + \mu dt$

2. *Geometric BM :* As our second example, consider geometric BM of the form

   $S(t) = S_0 e^{\sigma B(t) + \mu t}$, where $S_0 > 0$ is a constant initial value.

   We use $f(t,x) = S_0 e^{\sigma x + \mu t}$ and observing that $\frac{\partial f}{\partial x} = \sigma f$, $\frac{\partial f}{\partial t} = \mu f$ and $\frac{\partial^2 f}{\partial x^2} = \sigma^2 f$, obtain

$$dS(t) = \sigma S(t)dB(t) + (\mu + \frac{1}{2}\sigma^2)S(t)dt.$$

This is the classic stochastic differential equation (SDE) for geometric BM.

Recalling that (via moment generating functions of normal rvs)

$$E(S(t)) = S_0 e^{\overline{r}t},$$

where $\overline{r} = \mu + \sigma^2/2$, we see how this constant $\overline{r}$ shows up in the SDE.

3. *Ornstein-Uhlenbeck process:*

Here we start with a SDE to define a process ($\alpha > 0$ and $\sigma > 0$ constants):

$$dX(t) = -\alpha X(t)dt + \sigma dB(t), \ X(0) = x_0,$$

which in integral form is $X(t) = x_0 - \alpha \int_0^t X(s)ds + \sigma B(t)$. Let's intuitively understand what such a process is like. Note the similarity with BM with drift, but now instead of a constant drift μ, the "drift" term $-\alpha X(t)dt$, is proportional to the current position of the process: when $X(t)$ is large and positive, the drift is large and negative, pulling the process back towards the origin, and similarly, when $X(t)$ is large and negative, the drift is large and positive yet again pulling the process back towards the origin. Otherwise, the process, due to its BM piece, moves like a BM.

Thus we expect such a process to behave very much like a BM that is forced to stay reasonably close to the origin.

It turns out that the solution to this SDE is given by

$$X(t) = x_0 e^{-\alpha t} + \sigma \int_0^t e^{-\alpha(t-s)} dB(s),$$

and is called the *Ornstein-Uhlenbeck* process. As $t \to \infty$ we note that the initial value x_0 wears off (and does so exponentially fast). The strong forcing towards the origin turns out to result in a steady-state distribution being reached as $t \to \infty$. To see what this limiting distribution might be, note that $\int_0^t e^{-\alpha(t-s)} dB(s)$ is a mean 0 martingale (because it is an Ito integral), hence $E(X(t)) = x_0 e^{-\alpha t} \to 0$. Moreover, the variance of $\sigma \int_0^t e^{-\alpha(t-s)} dB(s)$ (having mean 0) equals its second moment which we know is the same as $\sigma^2 \int_0^t e^{-2\alpha(t-s)} ds = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t})$ via the Ito isometry. Thus $Var(X(t)) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) \to \frac{\sigma^2}{2\alpha}$ and we conclude that the limiting distribution should have mean 0 and variance $\frac{\sigma^2}{2\alpha}$. But what should the distribution be? The missing link is the fact that $\int_0^t e^{-\alpha(t-s)} dB(s)$ is a Gaussian process since the integrand is deterministic; this is a consequence of the fact that weighted sums of independent normal rvs are again normal; the integral is such a weighted sum. Thus the Ornstein-Uhlenbeck process $\{X(t)\}$ is a Gaussian process and its limiting distribution is $N(0, \frac{\sigma^2}{2\alpha})$.

On HMWK 4, you may recall: For $c > 0$, the Gaussian process $U(t) = e^{-ct/2} B(ce^{ct})$, has the same distribution for all $t \geq 0$; it is $N(0, c)$. If we let $c = \frac{\sigma^2}{2\alpha}$, then we have a stationary Gaussian stochastic process with our desired distribution; that this is an explicit construction of a stationary version of our Ornstein-Uhlenbeck process is what you will show in HMWK 6.

1.2 Generating martingales as functions of BM

Proposition 1.1 *If $f = f(t,x)$ is a function such that both $\frac{\partial f}{\partial t}(t,x)$ and $\frac{\partial^2 f}{\partial x^2}(t,x)$ are continuous, and if*

$\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0$, *then $M(t) = f(t, B(t)),\ t \geq 0$ is a martingale if in addition $X(t) = \frac{\partial f}{\partial x}(t, B(t))$ forms a square integrable process.*

Proof :

The key point is that the condition $\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0$ implies (via Ito's formula) that $M(t) = f(0,0) + \int_0^t X(s)dB(s)$, e.g., that M is an Ito integral (with mean $f(0,0)$). ■

Application to geometric BM

Using $f(t,x) = S_0 e^{\sigma x + \mu t}$ yields geometric BM, $S(t) = S_0 e^{\sigma B(t) + \mu t}$. As derived above, Ito's formula yields the SDE

$$dS(t) = \sigma S(t)dB(t) + (\mu + \frac{1}{2}\sigma^2)dt.$$

$\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0$ iff $\mu = -\frac{1}{2}\sigma^2$, yielding the SDE $dS(t) = \sigma S(t)dB(t)$, corresponding to a geometric BM of the special form

$$S(t) = S_0 e^{\sigma B(t) - \frac{1}{2}\sigma^2 t};$$

thus we conclude that the function $f(t,x) = S_0 e^{\sigma x - \frac{1}{2}\sigma^2 t}$ always yields a martingale for any $\sigma > 0$. (It is easily verified that $\{S(t)\}$ is a square integrable process, given as a HMWK set 6 problem.)

In the context of modeling stock pricing, there is the interest rate $r > 0$ to incorporate, in which case the discounted stock price per share is of the form $D(t) = S_0 e^{-rt} e^{\sigma B(t) + \mu t}$; the idea being that we want to consider the present value of the future price as opposed to the price. This can be obtained from Ito's formula via using the function $f(t,x) = S_0 e^{\sigma x + (\mu - r)t}$. In this case, $D(t)$ forms a MG if $\mu = r - \frac{1}{2}\sigma^2$. (In practice $r >> \sigma^2$ so this drift parameter μ would be positive.) This special choice of μ turns out to be of fundamental importance in option pricing; we will learn more about this later.

1.3 Ito processes

Ito's formula leads to stochastic processes with continuous sample paths of the form

$$Y(t) = Y(0) + \int_0^t K(s)ds + \int_0^t H(s)dB(s), \tag{1}$$

where K and H are stochastic processes adapted to the underlying filtration $\mathcal{F}_t,\ t \geq 0$, and $Y(0) \in \mathcal{F}_0$. For example, using Ito directly on $Y(t) = f(t, B(t))$ is the special case when $K(t) = \frac{\partial f}{\partial t}(t, B(t))$, and $H(t) = \frac{\partial f}{\partial x}(t, B(t))$. Allowing K and H to be general adapted processes yields *Ito processes* when we make the further assumptions:

A1 wp1,

$$\int_0^t |K(s)|ds < \infty,\ t \geq 0.$$

A2 wp1,

$$\int_0^t H^2(s)ds < \infty,\ t \geq 0.$$

Condition **A2** is weaker than the square integrability Condition *4*, $E\big[\int_0^t H^2(s)ds\big] < \infty,\ t \geq 0$, from our framework that we assumed originally for defining the Ito integral, but it turns out that it is enough to define the Ito integral (we leave out the technical details here), the integral can be extended to cover these more general processes. The main difference is that this weaker condition does not ensure that the Ito integral yields a martingale (it does however yields what is called a *local* martingale), and of course the Ito isometry makes no sense unless Condition *4* holds. But we will not dwell on this.

The differential form of (1) is given by

$$dY(t) = K(t)dt + H(t)dB(t), \tag{2}$$

which tells us that we can now define a stochastic integral with respect to Y via

$$\int_0^t X(s)dY(s) \stackrel{\text{def}}{=} \int_0^t X(s)K(t)dt + \int_0^t X(s)H(s)dB(s), \tag{3}$$

thus reducing the stochastic integral to regular dt integration plus a regular Ito integral.

To define such integrals against dY we could have started from scratch by defining the stochastic integral as we did for the Ito integral, replacing B by Y; but we would get the same thing in the end.

1.4 Ito's formula for Ito processes and integration by parts

Theorem 1.2 *Suppose $Y(t) = Y(0) + \int_0^t K(s)ds + \int_0^t H(s)dB(s)$ is an Ito process. If $f = f(t,x)$ is a function such that both $\frac{\partial f}{\partial t}(t,x)$ and $\frac{\partial^2 f}{\partial x^2}(t,x)$ are continuous, then*

$$f(t,Y(t)) = f(0,Y(0)) + \int_0^t \frac{\partial f}{\partial x}(s,Y(s))dY(s) + \int_0^t \frac{\partial f}{\partial t}(s,Y(s))ds + \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(s,Y(s))d < Y,Y >_s,$$

where

$$< Y,Y >_t \stackrel{\text{def}}{=} \int_0^t H^2(s)ds,$$

the quadratic variation of Y, and (recall (3))

$$\int_0^t \frac{\partial f}{\partial x}(s,Y(s))dY(s) = \int_0^t \frac{\partial f}{\partial x}(s,Y(s))K(s)ds + \int_0^t \frac{\partial f}{\partial x}(s,Y(s))H(s)dB(s).$$

In differential form: Let $Z(t) = f(t, Y(t))$. Then

$$\begin{aligned} dZ(t) &= \frac{\partial f}{\partial x} dY(t) + \frac{\partial f}{\partial t} dt + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} d < Y, Y >_t \\ &= \Big[\frac{\partial f}{\partial x} K(t) + \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} H^2(t) \Big] dt + \frac{\partial f}{\partial x} H(t) dB(t). \end{aligned}$$

Proposition 1.2 (Integration by parts) *Suppose $U(t) = U(0) + \int_0^t K_U(s)ds + \int_0^t H_U(s)dB(s)$ is an Ito process, and $V(t) = V(0) + \int_0^t K_V(s)ds + \int_0^t H_V(s)dB(s)$ is another. Then*

$$\int_0^t U(s)dV(s) = U(t)V(t) - U(0)V(0) - \int_0^t V(s)dU(s) - < U, V >_t,$$

where

$$< U, V >_t \stackrel{\text{def}}{=} \int_0^t H_U(s) H_V(s) ds.$$

In differential form this yields the product rule:

$$d(UV) = VdU + UdV + d < U, V >_t .$$

Proof : Using Ito's formula separately on each of the three processes, $(U+V)^2$, U^2, V^2, yields

$$\begin{aligned} (U(t) + V(t))^2 &= (U(0) + V(0))^2 + 2 \int_0^t (U(s) + V(s)) d(U(s) + V(s)) + \int_0^t (H_U(s) + H_V(s))^2 ds \\ U^2(t) &= U^2(0) + 2 \int_0^t U(s) dU(s) + \int_0^t H_U^2(s) ds \\ V^2(t) &= V^2(0) + 2 \int_0^t V(s) dV(s) + \int_0^t H_V^2(s) ds. \end{aligned}$$

subtracting the second two from the first yields the result.

■

We mention in passing that $< U, V >_t$ is called the *cross variation* of U and V over $[0, t]$ and can be identified as (via taking partitions Π of $[0, t]$)

$$\lim_{|\Pi| \to 0} \sum_{k=1}^{n} |U(t_{k-1}) - U(t_k)| |V(t_{k-1}) - V(t_k)|.$$

1.5 Stochastic differential equations

We have now seen that Ito's formula leads to "differential" equations of the form

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dB(t), \ X(0) = x_0 \qquad (4)$$

where $\mu = \mu(t,x)$ and $\sigma = \sigma(t,x)$ are deterministic functions. We view μ as the *drift* function and σ the *variability* function. Effectively we are generalizing a BM with constant drift μ and constant variance parameter σ to allow for time changing parameters that also depend on the state of the process itself at any given time. We call (4) a stochastic differential equation (SDE).

For example, geometric BM yields the SDE $dX(t) = (\mu + \sigma^2/2)X(t)dt + \sigma X(t)dB(t)$ which is the case when $\mu(t,x) = (\mu + \sigma^2/2)x$ (a linear function in x), and $\sigma(t,x) = \sigma x$ (a linear function in x).

We also saw how we came up with the Ornstein-Uhlenbeck process by at first considering the SDE $dX(t) = -\alpha X(t)dt + \sigma dB(t)$, which is the case when $\mu(t,x) = -\alpha x$, and $\sigma(t,x) = \sigma$ (a constant).

In integral form (4) becomes

$$X(t) = x_0 + \int_0^t \mu(s, X(s))ds + \int_0^t \sigma(s, X(s))dB(s).$$

Processes satisfying (4) are called *Ito diffusions* via viewing them as describing the position at time t of a particle suspended in a moving fluid under random molecular bombardment. (This is how Brownian motion was originally explained (by Einstein) via molecular bombardment by water molecules (of a pollen particle).)

It is natural to consider in general when (4) has a solution. This amounts to finding the kinds of functions μ and σ allowed, that is, for which a solution can be proven to exist. It turns out that reasonable conditions have been found. To be precise, the following are sufficient conditions: For every fixed $t > 0$, there exists constants $C > 0$ and $D > 0$ (depending on t) such that for all x,y and all $s \in [0,t]$,

$$|\mu(s,x)| + |\sigma(s,x)| \leq C(1 + |x|)$$

and

$$|\mu(s,x) - \mu(s,y)| + |\sigma(s,x) - \sigma(s,y)| \leq D|x - y|.$$

The second condition is a so-called Lipschitz condition implying that both functions must be Lipschitz in the x variable for each fixed $s \in [0,t]$. (This condition ensures uniqueness of solutions.) The first condition is a growth condition implying essentially that for each fixed s, the functions can not grow faster than linear in x. (Without this condition, it would be possible to make a process *explode* meaning that it would hit the value ∞ in finite time.)

Under these restrictions, (4) has a unique (wp1) solution $\{X(t) : t \geq 0\}$ that has continuous sample paths and is square integrable. Moroever, the solution is adapted to the BM: For $\mathcal{F}_t = \sigma\{B(s) : 0 \leq s \leq t\}$, it holds that $X(t) \in \mathcal{F}_t,\ t \geq 0$. (The initial state x_0 is allowed to be a random variable independent of the BM, in which case it must be thrown in to the filtration to form $\mathcal{F}_t^{x_0} = \sigma\{B(s) : 0 \leq s \leq t,\ x_0\}$.)

Time-homogenous Markov processes

In the examples above we note that the drift and variability functions did not depend on t, and this more generally yields a nice special case of an SDE of the form

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dB(t),\ X(0) = x_0, \tag{5}$$

where $\mu = \mu(x)$ and $\sigma = \sigma(x)$. A process satisfying (5) is called a *time-homogenous* Ito diffusion, and such processes are always Markov processes. The Markov property is staring at us in (5): If the present state $X(t)$ is given, then the future (via a future increment $dX(t)$) only depends additionally on a future increment $dB(t)$ of the BM which of course is independent of the past (e.g., independent of $\mathcal{F}_t$).

In particular we conclude that both geometric BM and the Ornstein-Uhlenbeck process are Markov processes.

We note in passing that the two sufficient conditions ensuring a solution in this case reduce to one condition: the existence of a $D > 0$ such that for all x, y

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq D|x - y|.$$

We finally note that in general, Ito diffusions are time-dependent Markov processes, meaning that the transition probabilities depend on what time it is as well as what the current state is.

1.6 APPENDIX

1.6.1 Doob's maximal inequalities

Earlier, in discrete time, we have seen one form of Doob's maximal inequality which we will present first here in continuous time, an then refine and expand it as well.

Proposition 1.3 (Doob's maximal I) *Let $\{X(t)\}$ be a non-negative submartingale, and let $M(t) \stackrel{\text{def}}{=} \max_{0 \leq s \leq t} X(s)$. Then $\forall t \geq 0$*

1.
$$P(M(t) > x) \leq \frac{E(X(t))}{x}, \; x \geq 0$$

2. *If in addition $E(X^2(t)) < \infty$, then*
$$P(M(t) > x) \leq \frac{E(X^2(t))}{x^2}, \; x \geq 0$$

Proof : For fixed t and x, define stopping time $\tau = \min\{0 \leq s \leq t : \; X(s) > x\}$ where $\tau \stackrel{\text{def}}{=} t$ if $X(s) \leq x, \; 0 \leq s \leq t$. Then $M(t) > x$ if and only if $X(\tau) > x$ and so Markov's inequality yields $P(M(t) > x) = P(X(\tau) > x) \leq E(X(\tau))/x$. Letting t be our second stopping time, both stopping times are bounded (by t) and $\tau \leq t$; thus the optional stopping theorem (Theorem 1.6 in Lecture Notes 2, for example) applies yielding $E(X(\tau)) \leq E(X(t))$, and the proof of *1* is complete. *2* follows by noting that if X is a non-negative submartingale, then so is X^2 as long as $E(X^2(t)) < \infty, \; t \geq 0$. Thus the result follows by noting that $\{M(t) > x\} = \{\max_{0 \leq s \leq t} X^2(s) > x^2\}$, and then applying *1* to X^2.

■

The following follows directly from the above proposition since $|X|$ is always a non-negative submartinagle if X is a martingale.

Corollary 1.1 *If $\{X(t)\}$ is a martingale, then*

1.

$$P(\max_{0\le s\le t}|X(s)|>x)\le \frac{E(|X(t)|)}{x},\ x\ge 0.$$

2. If in addition $E(X^2(t))<\infty$, then

$$P(\max_{0\le s\le t}|X(s)|>x)\le \frac{E(X^2(t))}{x^2},\ x\ge 0.$$

Now we refine Doob's inequalities in discrete time so as to then do so in continuous time:

Proposition 1.4 (Doob's maximal II in discrete time) *Let $\{X_n\}$ be a non-negative submartingale, and let $M_n \stackrel{\text{def}}{=} \max_{0\le k\le n} X_k$. Then $\forall n\ge 0$*

1.

$$xP(M_n>x)\le E(X_nI\{M_n>x\})\le E(X_n),\ x\ge 0. \tag{6}$$

2. If in addition $E(X_n^2)<\infty$, then

$$E(M_n^2)\le 4E(X_n^2), \tag{7}$$

in particular $E(M_n^2)<\infty$.

Proof : Let $\tau=\min\{0\le k\le n: M_k>x\}$, $\tau\stackrel{\text{def}}{=} n$ if $M_k<x,\ 0\le k\le n$. Then $I\{M_n>x\}=I\{X_\tau>x\}$ and since $I^2=I$ for any indicator, $I\{M_n>x\}=I\{X_\tau>x\}I\{M_n>x\}$. Observing that $xI\{X_\tau>x\}\le X_\tau$ yields

$$xI\{M_n>x\}=xI\{X_\tau>x\}I\{M_n>x\}\le X_\tau I\{M_n>x\},$$

and thus

$$xP(M_n>x)\le E(X_\tau\{M_n>x\}). \tag{8}$$

But for $0\le k\le n$,

$$\begin{aligned}
E(X_\tau I\{M_n>x\}I\{\tau=k\}) &= E(X_kI\{M_k>x\}I\{\tau=k\}) && (9)\\
&\le E(E(X_n\mid \mathcal{F}_k)I\{M_k>x\}I\{\tau=k\})\ \text{(subMG property)} && (10)\\
&= E(E(X_nI\{M_k>x\}I\{\tau=k\}\mid \mathcal{F}_k) && (11)\\
&= E(X_nI\{M_k>x\}I\{\tau=k\}) && (12)\\
&= E(X_nI\{M_n>x\}I\{\tau=k\}), && (13)
\end{aligned}$$

thus summing up over k, $0\le k\le n$, yields $E(X_\tau I\{M_n>x\})\le E(X_nI\{M_n>x\})$ which when plugged into (8) yields (6); the proof of *1* is complete. For *2*: Recall that for any non-negative rv Z, its moments can be computed via integrating the tail; e.g.,

$E(Z) = \int_0^\infty P(Z > x)dx$, $E(Z^2) = \int_0^\infty 2xP(Z > x)dx$ and in general, for $p \geq 1$, $E(Z^p) = \int_0^\infty px^{p-1}P(Z > x)dx$. Applying this 2nd moment fact to M_n via (6) yields

$$\begin{aligned} E(M_n^2) &\leq 2\int_0^\infty E(X_n I\{M_n > x\})dx \\ &= 2E\Big[\int_0^\infty X_n I\{M_n > x\}dx\Big] \\ &= 2E\Big[\int_0^{M_n} X_n dx\Big] \\ &= 2E(X_n M_n). \end{aligned} \tag{14}$$

But actually, before we proceed further, we must be careful since we do not know apriori that $E(X_n M_n) < \infty$, all we know is that $E(X_n^2) < \infty$. To get around this, we first choose $a > 0$ and consider the truncated, hence bounded, rvs $M_n(a) = \min\{M_n, a\}$, and observe that the inequality (6) remains valid [1]:

$$xP(M_n(a) > x) \leq E(X_n I\{M_n(a) > x\}). \tag{15}$$

Thus re-doing what led us to (14) yields $E(M_n^2(a)) \leq 2E(X_n M_n(a)) \leq \sqrt{E(X_n^2)}\sqrt{E(M_n^2(a))}$, where for the last inequality we used Holder's inequality ($p = 2$), allowed since both X_n and $M_n(a)$ are in L^2. Dividing both sides by $\sqrt{E(M_n^2(a))}$ yields $\sqrt{E(M_n^2(a))} \leq 2\sqrt{E(X_n^2)}$ and then squaring both sides yields $E(M_n^2(a)) \leq 4E(X_n^2)$. Finally, applying Fatou's lemma as $a \to \infty$ yields the result. ■

Proposition 1.5 (Doob's maximal II) *Let $\{X(t)\}$ be a non-negative submartingale with continuous sample paths, and let $M(t) \stackrel{\text{def}}{=} \max_{0 \leq s \leq t} X(s)$. Then $\forall t \geq 0$*

1. $$xP(M(t) > x) \leq E(X(t)I\{M(t) > x\}),\ x \geq 0$$

2. *If in addition $E(X^2(t)) < \infty$, then*

   $$E(M^2(t)) \leq 4E(X^2(t)),$$

   in particular $E(M^2(t)) < \infty$.

Proof : We will discretize the problem and use Proposition 1.4: Partition $[0, t]$ into the n equal length subintervals (of length t/n) via $t_k = kt/n$, $0 \leq k \leq n$, and consider the discrete-time submartingale $X_k = X(t_k)$ on $\{0 \leq k \leq n\}$, with its maximum $M_n = \max_{0 \leq k \leq n} X_k$. By the assumed continuity of sample paths,

$$\lim_{n \to \infty} M_n = M(t),\ wp1, \tag{16}$$

[1] To see this, note that if $a \geq x$, then $I\{M_n(a) > x\} = 0$ and so $P(M_n(a) > x) = 0$ and thus both sides of (15) are 0; whereas if $a < x$, then $I\{M_n(a) > x\} = I\{M_n > x\}$ and both sides of (15) are as in *1.*

and so $\forall x \geq 0$,

$$\lim_{n\to\infty} I\{M_n > x\} = \lim_{n\to\infty} \{M(t) > x\},\ wp1. \tag{17}$$

Now, since $X_n = X(t)$, (6) becomes

$$xP(M_n > x) \leq E(X(t)I\{M_n > x\}) \leq E(X(t)),\ x \geq 0, \tag{18}$$

and the dominated convergence theorem can be used on both $P(M_n > x)$ and $E(X(t)I\{M_n > x\})$ yielding *1*. The proof of *2* is identical to that in Proposition 1.4, using the same truncation argument.

∎