

4. Stochastic Calculus-1

0. ODE(Ordinary Differentiation Equation)

현실에서 어떠한 현상의 함수를 알아내는 것은 무척 어려운 일이다. 보통은 어떠한 대상의 변화율과 함숫값, 시간정보만이 주어지고, 그로 부터 나온 변화율의 함수인 미분 방정식을 이용해 원래의 함수를 추정해내야 한다.

ODE가 사용될 수 있는 계산의 상황은 '결정론적인' 상황이다. 시간 t의 값만 알면 미래의 함수값도 쉽게 알아낼 수 있는 이상적인 상황이다.

$$\frac{df}{dt} = f'(t)$$
 $df(t) = f'(t)dt$

하지만 확률에 의해 표현되는 확률과정을 ODE를 이용해 표현할 수 있을까?

1. SDE (Stochastic Differentiation Equation)

2. Wiener Process에서 다뤘다시피, 확률과정은 다음과 같이 표현 가능하다.

$$X(t) = \mu t + \sigma W(t)$$

정교화를 위해 각각에 곱해져있는 계수들을 함수로 정의하면, 다음과 같다.

$$X(t) = \alpha(t, X_t)t + \sigma(t, X_t)W(t)$$

그럼 ODE를 이용해 확률과정의 미분 방정식을 세운다면 다음과 같을 것이다.

$$egin{aligned} \Delta X(t) &= lpha(t,X_t)\Delta t + \sigma(t,X_t)\Delta W(t) \ dX(t) &= lpha(t,X_t)dt + \sigma(t,X_t)dW(t) \ \dfrac{dX(t)}{dt} &= lpha(t,X_t) + \sigma(t,X_t)\dfrac{dW(t)}{dt} \end{aligned}$$

하지만 앞서 위너 프로세스의 미분 불가능성을 직관적인 그림으로 확인한 바가 있다. 따라서 제일 오른쪽의 항(위너가 있는 항) 때문에 이 공식은 해를 찾을 수 없다.

<Quadratic Variation of Wiener>



 $\frac{dW(t)}{t}$ 를 다루는 문제를 피하기 위해, 우리는 위 공식의 적분형을 다룰 것이다.

$$dX(t) = lpha(t,X_t)tdt + \sigma(t,X_t)dW(t) \ X_t = X_0 + \int_0^t lpha\left(s,X_s
ight)ds + \int_0^t \sigma\left(s,X_s
ight)dW_s$$

이로써 W(t)가 t와 엮이는 일이 없이, W(t)를 변수로 하는 새로운 적분이 뒤에 정의된다. t에 대해 직접적으로 미분이 되는 것을 피하는 일 종의 우회로를 택한 것이다. 따라서 W(t)를 포함하는 SDE는 이름만 확률미분방정식이지 그 형태는 적분방정식이다.

1.1. SDE형의 의미

SDE형에 대한 이해는 다음의 경우로 가능하다.

우리가 FR식당의 주인이고, 확률 과정으로 가게의 총 매출을 예측하고 싶다. 우선은 다음의 논리로 단골 손님의 수를 표현하자. 단골 손님 은, FR식당에 갈 마음이 들 확률이 90%인 사람들이라고 가정하자. 그렇다면 그들의 방문 여부는 다음의 확률 분포로 결정될 것이다.

명수	0명	1명	2명	 t-1명	t명	
확률	$t\mathcal{C}_0(0.1)^t$	$tC_1(0.1)^{t-1}(0.9)^1$	$tC_2(0.1)^{t-2}(0.9)^2$	$tC_{t-1}(0.1)^1(0.9)^{t-1}$	$tC_t(0.9)^t$	

t의 매 순간마다, 식당에 방문하는 경우를 v, 방문하지 않는 경우를 f라 하면,

0명이 방문하는 경우 = (F, F, F, ...F)

1명이 방문하는 경우 = (V, F, F, ...F)(F, V, F, ...F), ..(F, F, F, ...V)

2명이 방문하는 경우 = (V, V, F, ...F)(F, V, V, ...F), ...(F, F, F, ...V, V)

t명이방문하는경우(V, V, V, ...V)

총 2^t 개의 시행 경우가 나올 것이다. 이러한 시행의 경우들을 각각 표본 ω 라고 해, 다음의 표본공간을 정의하자.

$$\Omega = \{\omega, \omega_1, \omega_2, \ldots, \omega_{2^t-1}, \omega_{2^t}\}$$

이제 위의 표본공간에서 변수를 받고, 시점 s의 영향을 받는 이변수 함수 $x(\omega,s)$ 를 정의할 수 있다. 이 함수는 s 시점에서, 단골 손님의 수라고 할 수 있을 것이다.

$$X_t = X_0 + \int_0^t lpha(\omega,s)x(\omega,s)ds + \int_0^t \sigma\left(\omega,s
ight)dN(W_s)$$

 $X_t -$ 촉매축

 $x(\omega, s) = s$ 시점에서 단골 손님 수

 $lpha(\omega,s)=$ 단골 손님이 시키는 메뉴의 가격

 $N(W_s) = s$ 시점에서 비정기적인 손님 수

 $\sigma\left(\omega,s
ight)=$ 비정기적으로 방문한 손님이 시키는 메뉴의 가격

이것은 FR식당의 총 매출을 나타내는 SDE가 되는 것이다. 이때, 우리는 현실에 맞게 단골 손님의 수가 어느 정도 랜덤한 확률 과정을 따른다고 가정했다. 하지만, 더욱 단순하게 단골 손님이 매 시점마다 5명씩 온다고 가정해 랜덤성을 완전히 제거한다면, 아래와 같이 랜덤한 부분과 랜덤하지 않은 부분으로 나눌 수 있을 것이다.

$$X_t = X_0 + \int_0^t 5 \; lpha(\omega,s) ds + \int_0^t \sigma \left(\omega,s
ight) dN(W_s)$$

위에서 살핀 바와 같이, 일반적인 주가 SDE 표현형은 다음과 같다. α 와 σ 는 시점 s에서의 추세와 변동성의 함수를 의미한다. 둘은 동일한식이다.

$$X_{t}=X_{0}+\int_{0}^{t}lpha\left(s,X_{s}
ight) ds+\int_{0}^{t}\sigma\left(s,X_{s}
ight) dW_{s}$$

블랙 숄즈 모델에선 α 와 σ 는 각각 기대수익률과 수익률의 변동성이다.

1.2. SDE에 여전히 들어있는 W(s)를 처리할 수 있는가?

Ito Integral을 이용하면 된다.

2. Itô calculus

위너 과정 W(t)의 미분 불가능성은 이후의 계산에 큰 난관이 된다. 아래의 예시를 보자. 함수 t는 위너 프로세스를 정의역으로 가지는 일종의 옵션 가격 결정 함수이다.

$$f(W(t)), f \in C^1$$

이러한 f의 시간에 대한 변화율이 궁금해 미분을 하면 다음과 같이 나온다.

$$rac{df(W(t))}{dt} = rac{df(W(t))}{dW(t)} imes rac{dW(t)}{dt}$$

하지만 뒤의 위너 과정은 미분이 불가능하기 때문에 우리는 변화율을 구할 수 없게 된다.

따라서 다른 방법을 시도해, f의 W(t)에 대한 변화율을 구한다고 해보자. 파생상품도 위너 프로세스를 변수로 받기에 이런 특성을 가지고 있다.

$$\frac{df(W(t))}{dW(t)} = f'(W(t))$$
$$df(W(t)) = f'(W(t)) \times dW(t)$$

이로써 아주 작은 시간 간격 동안 W(t)의 변화에 대한 f의 변화율을 표현할 수 있게 된다.

하지만 W(t)는 위의 표현형 또한 옳지 않다. 전통적인 미분의 상황은 테일러 급수를 이용해 다음과 같이 표현할 수 있다.

$$f(t+x) = f(t) + f'(t)x + \frac{f''(t)}{2}x^2 + \frac{f'''(t)}{3!}x^3 + \dots$$

$$f(t+x) - f(t) = f'(t)x + \frac{f''(t)}{2}x^2 + \frac{f'''(t)}{3!}x^3 + \dots$$

$$\lim_{x \to 0} \frac{f(t+x) - f(t)}{x} = \frac{df}{dt} = f'(t)$$
so
$$\frac{df(W(t))}{dW(t)} = f'(W(t))$$
?

x 는 시간 증분이며, x가 0으로 가기 때문에 뒤의 x^2, x^3등을 무시해줄 수 있게 된다.

하지만 $\frac{d(f(W(t)))}{W(t)}$ 의 경우처럼, 변수를 W(t)로 받는다고 생각하면 문제가 생긴다. 따라서 앞선 단원에서 Quadratic Variation의 여러 특성들 중, dWdW=dt의 결과를 이용해 고려해보자.

$$f(W_{t+x}) = f(W_t) + f'(W_t)(W_{t+x} - W_t) + \frac{f''(W_t)}{2}(W_{t+x} - W_t)^2 + \frac{f'''(W_t)}{3!}(W_{t+x} - W_t)^3 + \dots$$

$$f(W_{t+x}) - f(W_t) = \Delta f(W), \ W_{t+x} - W_t = \Delta W$$

$$\Delta f(W) = f'(W_t)\Delta W + \frac{f''(W_t)}{2}\Delta W^2 + \frac{f'''(W_t)}{3!}\Delta W^3 + \dots$$

$$||\pi|| \to 0, \ \Delta f(W) \to df, \ \Delta W \to dW$$

$$df = f'(W_t)dW + \frac{f''(W_t)}{2}dW^2 + \frac{f'''(W_t)}{3!}dW^3 + \dots$$

$$df = f'(W_t)dW_t + \frac{f''(W_t)}{2}dt$$

$$*dWdW = dt, \ dWdt = 0, \ dtdt = 0$$

앞서 W(t)를 변수로 f의 변화량을 측정한다고 세운 식과, dt가 나타난 새로운 식을 비교해보자. 시간을 무한히 쪼갤때, 일반적인 미분의 상황에서는 x가 빠른 속도로 극한값으로 수렴한다.



하지만 위너의 경우에선, 테일러 전개에서 튀어나온 dt의 크기를 우리는 무시할 수 없다. 반대로 적분을 해준다고 하면, $(W_{t+x}-W_t)^2$ 은 T로 나오게 된다. 그렇기에 W(t)는 각각 일반적인 미분과 적분이 불가능하다.

아래의 식이 W(t)를 다루기 위한 특수한 미적분 체계인 Ito calculus의 직관을 담고 있다. 깊이 들어가면, Ito calculus 일반적인 미적분학 과의 가정과는 다른 수많은 가정들을 다뤄야 한다. 하지만 우리는 Ito calculus가 확률과정의 2차 변동성까지 고려하는, 일종의 '섬세한' 미분법이라고 이해하고 넘어가고 그 실질적인 적용에 집중할 것이다. 또한, 미분이 시간 t에 대해서가 아니라, W(t)에 대해서 이뤄진다는 것을 기억하자.

$$X_{t}=X_{0}+\int_{0}^{t}lpha\left(s,X_{s}
ight) ds+\int_{0}^{t}\sigma\left(s,X_{s}
ight) dW_{s}$$

앞에서 간단하게 정의한 f함수가 단순한 일반적인 함수가 아닌 확률과정이라면 훨씬 복잡한 일반항을 얻게 된다. 우리는 이러한 직관으로 부터 수많은 미적 법칙이 파생된다는 것만 알고, 기술적인 적용과 특성들만 아는 수준으로 넘어간다.

3. Stochastic Itô Integral

우리가 배운 일반적인 적분 과정은 다음과 같이 변수 t를 매개로 적분한다.

$$\sum_{n=0}^{N-1} f_{t_n} \Delta t
ightarrow \int_0^T f_t dt$$

하지만 이토 적분 과정은 앞서 보인듯이, Wt를 변수로 사용한다. 따라서 다음의 형태로 이뤄진다.

$$\sum_{n=0}^{N-1} f_{t_n} \Delta W_t
ightarrow \int_0^T f_t dW_t$$

4. Properties of Itô Integral

이토 적분의 정확한 적용에 들어가기 전, 이토 적분의 특성들을 살펴보자.

과정 X를 adapted한 확률 과정이라고 하자. 즉, 각 실현 사건은 Filtration에 속하며, 그에 따라 미래의 기댓값은 현재와 동일한 fair game 이다. 이때, X는 W(s)에서 주는 정보를 바탕으로 adapted 된다.

$$I(t) = \int_0^t X(s)dW_s$$

W(s)가 어떠한 노름의 결과라고 하면, 그에 곱해지는 과정 X는 노름꾼의 베팅 전략이라고 할 수 있다.

또한 앞선 단원에서 복잡하게 표현됐지만, 결국 adapted 됐다는 말은, 노름꾼이 미래를 내다볼 수 없다는 현실의 수학적 표현형이다. 그에 따라 노름꾼이 자신의 베팅 전략을 수정할 때 참고할 정보는, 오직 현재의 노름의 결과인 W(s)뿐이다.

할 수 있다. 가지고 있는 주식의 수량 또한 전략의 일종이고, 주식의 가격에 따라 결정하기 때문이다. 따라서 I(t)는 Profit/Loss가 된다.

금융의 경우, 만일 과정 X와 곱해지는 과정이 위너 프로세스를 따르는 주식의 가격이라 한다면, 그에 곱해지는 X는 주식의 수와 같다고

$$\mathbb{E}\left[\int_0^t X^2(s)ds
ight] < \infty$$
 , then $I(t) = \int_0^t X(s)dW(s)$ has the following properties

X 제곱을 적분해 기댓값을 취했을 때, 발산하지 않으면 W(s)에 대한 X 의 이토 적분을 정의할 수 있다. 아래의 식에서는 B를 f(s,w)로 표현해. 이토를 정의했다.

$$\int_{0}^{T}f(s,w)dW_{s}(w)\ where\ f:[0,T] imes\Omega o\mathbb{R}, orall\ W\in\Omega$$

이 조건이 만족하면 이토 적분은 다음과 같은 특성들을 지닌다.

4.1 Continuity

I(t)는 연속함수이다.

4.2 Adaptivity

각 t에 대해 I(t)는 ${\cal F}$ -measurable하다. 즉, adapted한 과정의 곱인 I(t) 또한 adaptivity를 지닌다.

4.3 Linearity

적분 가능하기 때문에 일반적인 적분의 연산법칙을 전부 적용할 수 있다.

$$\int_0^t Y(u)dW(u) \pm c \int_0^t \Gamma(u)dW(u) = \int_0^t \Delta(u) \pm \int_0^t c \cdot \Gamma(u)dW(u)$$

4.4 I(t) is a martingale

$$\mathbb{E}[\xi_{n+1}] = \mathbb{E}[\xi_n] ext{ (by } 3.\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X])$$

마틴게일의 핵심적인 특성은 $E[\xi_{n+1}|\mathcal{F}_n]=\xi_n$ 로 정의되는 fair game이다. 이때, I(t)는 모든 t에서 적분 가능하며, I(t)를 이루는 두 확률과정 X와 W 모두가 W에 대해 measurable하기 때문에 미래의 기댓값은 전부 현재의 기댓값과 동일하게 된다. 시점 0에서의 I(t)는 0과 같기 때문에, 이후의 기댓값 또한 0이 된다.

$$E(I(0)) = E(I(t)) = 0$$

직관적으로 이해하자면, 주가가 만일 fair game을 따르고, 주식의 수량이 오직 주가에만 영향을 받는다고 했을 때, 둘의 곱으로 결정되는 함수인 수익 또한 미래의 기댓값과 현재의 기댓값이 동일하다는 것이다.

4.5 Itô Isometry

$$\begin{split} \mathbb{E}\left[I^2(t)\right] &= E\left(\left(\int_0^t X(u)dW(u)\right)^2\right) \\ &= E\left(\int_0^t \left(X(u)dW(u)\right)^2\right) \\ &= E\left(\int_0^t \left(X^2(u)du\right)\right) \\ &= E\left(\int_0^t X^2(u)\right)du \\ &= \int_0^t E\left(X^2(u)\right)du \\ &= \int_0^t E\left(X^2(u)\right)du \\ &* dWdW = dt, \ dWdt = 0, \ dtdt = 0 \\ &\texttt{t} \ \text{대신} \ \mathbf{u} \ \vec{=} \ \text{시간변수로} \ \Delta \text{면} \end{split}$$

dWdW = du

이로써 W가 포함된 식을 일반함수의 형태로 표현할 수 있게 된다.

또한, 이후에 Ito Isometry를 이용해 W의 방해 없이 위너 프로세스의 적분값을 편리하게 구해낼 수 있다.

4.6 Itô Isometry를 이용한 곱셈 법칙

만일 X와 동일한 조건을 따르는 두 과정 X와 G가 있다고 할 때, Ito Isometry를 이용해 다음과 같은 곱셈 법칙을 활용할 수 있다.

$$I(G) = \int_0^t G(u)dW(u), \ I(X) = \int_0^t X(u)dW(u)$$

$$\mathbb{E}\left[I(X)I(G)\right] = E(\int_0^t X(u)dW(u)) \int_0^t G(u)dW(u))$$

$$= E\left(\int_0^t (X(u)G(u)dW(u))^2\right)$$

$$= E\left(\int_0^t X(u)G(u)\right) du$$

$$= \int_0^t E\left(X(u)G(u)\right) du$$

4.7 Quadratic Variation of Itô Integral

앞서 QV를 확률과정의 변동성을 나타내는 단위 중의 하나로 브라운 운동이나 stochastic process를 분석할 때 쓰이는 개념으로 소개했다. 확률과정인 Ito 과정의 특성을 분석하기 위해 QV가 필요할 수도 있다.

$$egin{aligned} *d[W,W] &= dWdW \ d[I,I](t) &= dI(t)dI(t) \ &= X(t)dW(t)X(t)dW(t) \ &= X^2(t)dW(t)dW(t) \ &= X^2(t)dt \ \ &= I,I](t) &= \int_0^t X^2(u)du \end{aligned}$$

5. Itô formula (stochastic chain rule)

Let
$$Y(t)=U(t,X(t)), \quad U:[t_0,T] imes\Omega o\mathbb{R}$$
 $dX(t)=lpha\,(t_1x(t))\,dt+\sigma\,(t_1,x(t))\,dW(t)$
$$Y(t): 파생상품의가격$$
 $X(t):$ 기초자산의가격 $t:$ 시간 $\Omega:$ 모든 사건, $\mathbb{R}:$ 실수 전체

우리가 구하려는 파생상품의 가격은 시간과 기초 자산 가격의 함수이다. 따라서 시간 t와 기초 자산 X(t)를 변수로 받는 Y(t)라는 파생상품의 확률과정을 정의하고, 그때 Y(t)의 SDE 형태가 어떻게 되는지 확인한다. 결론은 다음과 같다.

Then,
$$dY(t) = L^0 U(t, X(t)) dt + L^1 U(t, X(t)) dW(t)$$
where
$$\begin{pmatrix} L^0 U = \frac{dU}{dt} + \alpha \cdot \frac{dU}{dx} + \frac{1}{2} \sigma^2 \frac{\partial^2 U}{dx^2} \\ L^1 U = \sigma \cdot \frac{\partial U}{\partial x} \end{pmatrix}$$

이러한 결론은 **Stochastic Chain Rule**을 통해 나왔다. 합성함수를 미분하는 방법인 Chain Rule은, ODE의 상황에서 다음과 같이 나온다. 하지만 이 Chain Rule은 불가능함을 이미 확인했다.

$$rac{df(W(t))}{dt} = rac{df(W(t))}{dW(t)} imes rac{dW(t)}{dt}$$

따라서 확률과정의 합성함수의 미분을 얻어내는 다른 방법이 바로 **Ito Formula**이다. 확률과정을 위한 합성함수 미분법이기 때문에, **Stochastic Chain Rule**이라고 불리기도 한다. 2개의 변수(시간, 기초자산)를 가지는 Y(t)(U(t, X(t))) 를 다변항의 테일러 전개를 이용해 미분한다.

다변항의 테일러 전개를 사용하는 이유는, 두 개의 변수를 받는 2차식을 미분해주기 위해서다.

$$f(x,y) - f(a,b) = \frac{\partial f}{\partial x}(x-a) + \frac{\partial f}{\partial y}(y-b) + \frac{1}{2}\frac{\partial^2 U}{\partial x^2}(x-a)^2 + \frac{\partial^2 U}{\partial x dy}(x-a)(y-b) + \frac{1}{2}\frac{\partial^2 U}{\partial y^2}(y-b)^2 + \dots$$

$$\text{cf})$$

$$\Delta X(t) = \alpha \cdot \Delta t + \sigma \cdot \Delta W(t)$$

$$(\Delta X(t))^2 = \alpha^2(\Delta t)^2 + 2\alpha\sigma\Delta t \cdot \Delta W(t) + \sigma^2(\Delta W(t))^2$$

$$\text{cf}) \quad dWdW = dt \quad / \quad dWdt = 0 \quad / \quad dtdt = 0$$

테일러 식을 이용해 Y(t)를 SDE형태로 도출하자.

$$f(x,y) - f(a,b) = \frac{\partial f}{\partial x}(x-a) + \frac{\partial f}{\partial y}(y-b) + \frac{1}{2}\frac{\partial^2 U}{\partial x^2}(x-a)^2 + \frac{\partial^2 U}{\partial x dy}(x-a)(y-b) + \frac{1}{2}\frac{\partial^2 U}{\partial y^2}(y-b)^2 + \dots$$

$$x=t+\Delta t, y=X(t)+\Delta X(t), a=t, b=X(t)$$
라고하면

$$\begin{split} \Delta Y(t) &= U(t + \Delta t, X(t) + \Delta X(t)) - U(t, X(t)) \\ &= \frac{\partial V}{\partial (t + \Delta t)} (t + \Delta t - t) + \frac{\partial U}{\partial (X + \Delta X)} (X(t) + \Delta X(t) - X(t)) \\ &+ \frac{1}{2} \frac{\partial^2 U}{\partial (t + \Delta t)^2} (\Delta t)^2 + \frac{\partial^2 U}{\partial (t + \Delta t) \partial (X + \Delta X)} \Delta (t) \Delta X(t) + \frac{\partial^2 U}{\partial (X + \Delta X)^2} (\Delta X(t))^2 \end{split}$$

분모의 $t+\Delta t$ 를 편의상 t로 표시하고, $X(t)+\Delta X(t)$ 를 편의상 X(t)로 표시한다.

$$\Delta Y(t) = U(t + \Delta t, X(t) + \Delta X(t)) - U(t, X(t))$$

$$= \frac{\partial V}{\partial t} \Delta t + \frac{dU}{\partial X} \Delta X(t) + \frac{1}{2} \frac{\partial^2 U}{\partial t^2} (\Delta t)^2 + \frac{\partial^2 U}{\partial t \partial X} \Delta t \Delta X(t) + \frac{1}{2} \frac{\partial^2 U}{\partial X^2} (\Delta X(t))^2 + \cdots$$

여기서 $\Delta X(t) = \alpha \cdot \Delta t + \sigma \cdot \Delta W(t)$ 를 대입한다.

$$\begin{split} &=\frac{\partial U}{\partial t}\Delta t + \frac{dU}{\partial X}(\alpha\cdot\Delta t + \sigma\cdot\Delta W(t)) + \frac{1}{2}\frac{\partial^2 U}{\partial t^2}(\Delta t)^2 + \frac{\partial^2 U}{\partial t\partial X}\Delta t(\alpha\cdot\Delta t + \sigma\cdot\Delta W(t)) + \frac{\partial^2 U}{\partial X^2}(\alpha\cdot\Delta t + \sigma\cdot\Delta W(t)) \\ &= \left(\frac{\partial U}{\partial t} + \alpha\cdot\frac{\partial U}{\partial X}\right)\Delta t + \sigma\frac{\partial U}{\partial X}\Delta W(t) + \left(\frac{1}{2}\frac{\partial^2 U}{\partial t^2} + \alpha\frac{\partial^2 U}{\partial t^2} + \frac{1}{2}\alpha^2\frac{\partial^2 U}{\partial X^2}\right)(\Delta t)^2 + \left(\sigma\frac{\partial^2 U}{\partial t\partial X} + \alpha\sigma\frac{\partial^2 U}{\partial X^2}\right)\Delta t\Delta W(t) \end{split}$$

여기서 dWdW=dt / dWdt=0 / dtdt=0 을 활용하여 항들을 정리한다. $=\Delta t$ 를 0으로 극한을 취당하여 함들을 정리한다.

$$=\left(rac{\partial U}{\partial t}+lpharac{\partial U}{\partial X}+rac{1}{2}\sigma^2rac{\partial^2 U}{\partial X^2}
ight)dt+\sigma\cdotrac{\partial U}{\partial X}dW(t)=L^0Udt+L^1UdW(t)$$

이로써 파생상품 확률과정의 일반항을 Stochastic Chain Rule을 이용해 얻어냈다.