Stochastic Calculus, Courant Institute, Fall 2020
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2020/index.html

Week 7

Jonathan Goodman, December, 2020

Tentative (more text)

1 Conditional expectation

Conditional expectation describes how our expected values change when we get partial information about the outcome. Fancy reasoning with conditional expectation expresses this differently from the elementary method using conditional probability densities or distributions. The

An example is the value function

$$f(x,t) = \mathrm{E}[\, V(X_T) \mid X_t = x] \ . \tag{1}$$

One formalism for this would involve the joint probability distribution

$$(X_t, X_T) \sim u(x_1, x_2) \ .$$

The conditional density of X_T conditional on $X_t = x_1$ is

$$u(x_2|x_1) = \frac{u(x_1,x_2)}{\int u(x_1,x_2)dx_1} \ .$$

The left side is a probability density as a function of x_2 for each value of x_1, but x_1 is just a parameter describing the partial information. Suppose $V(x_1, x_2)$ is a random variable whose value depends on X_1 and X_2. The conditional expectation of V, if X_1 is known, depends on the value of $X_1 = x_1$, but it is still a random variable. We call it $\widetilde{V}(x_1)$

$$\widetilde{V}(x_1) = \mathrm{E}[\, V(X_1,X_2) \mid X_1 = x_1\,] = \int V(x_1,x_2)\, u(x_2|x_1)\, dx_2 \ .$$

If we don't know the value of X_1, this conditional expectation is a random variable $\widetilde{V}(X_1)$. This is like $V(X_1, X_2)$, except that it depends only on X_1. The tower property gives

$$\mathrm{E}[\widetilde{V}(X_1)] = \mathrm{E}[V(X_1,X_2)] \ .$$

Partal information about a complex random variable X means learning the values of some components of X. In the example above, we learned one of the two components of X. The conditional expectation of $V(X)$ is the expectation given that information. When X is a path, partial information might mean knowing the first part of the path but not all of it. If the path is $X_{[0,T]}$ and $t < T$, we might know $X_{[0,t]}$. The conditional expectation would be the expectation knowing this information. Like the example, the conditional expectation is a function of what you know. So there could be a function $\widetilde{V}(x_{[0,t]})$ so that

$$\widetilde{V}(x_{[0,t]}) = \mathrm{E}[\, V(X_{[0,T]}) \mid X_{[0,t]} = x_{[0,t]}\,] \; .$$

Then there is a random variable that depends on the part of the path that you know $\widetilde{V}(X_{[0,t]})$. The tower property gives

$$\mathrm{E}[\, V(X_{[0,T]})\,] = \mathrm{E}[\, \widetilde{V}(X_{[0,t]})\,] \; .$$

As an example, suppose X_t is Brownian motion and consider the random variable

$$V(X_{[0,T]}) = \int_0^T e^s dX_s \; .$$

The conditional expectation, if you know $X_{[0,t]}$, is

$$\widetilde{V}(X_{[0,t]}) = \int_0^t e^s dX_s \; . \tag{2}$$

You can understand this claim by writing the whole integral in two parts

$$\int_0^t e^s dX_s + \int_t^T e^s dX_s \; .$$

When you know $X_{[0,t]}$, the first integral is known. The independent increments property says the second integral is independent of the first. Even conditional on $X_{[0,t]}$, the expected value of dX_s for $s > t$ is zero. Thus, the conditional expectation of the integral from t to T is zero.

As another example, consider a final time payout $V(X_T)$. The conditional expectation is given in terms of the value function

$$\mathrm{E}[\, V(X_T)\,] = f(X_t, t) \; . \tag{3}$$

This is because of the Markov property of X_t. The probability distribution of X_s for $s > t$ depends only on X_t and not on values X_s for $s < t$. The conditional expectation given $X_t = x$ is the value function.

Conditional expectations like this are often expressed using $\sigma-$*algebras* and *filtrations*. These are useful in rigorous mathematical discussions but seem to get in the way in less rigorous discussions. Nevertheless, you will see this terminology a lot if you practice stochastic calculus. A $\sigma-$ algebra is a collection of *events*. An event is a set of outcomes that is defined by some criterion or limiting procedure. For example, there is the event $A = \{|X_t| < 2 \text{ for all } t \leq 5\}$. An *algebra*, written $\mathcal{F}$, of events represents a state of partial information. An algebra of sets is a collection of events. Any event A either has $A \in \mathcal{F}$ or $A \notin \mathcal{F}$. If $A \in \mathcal{F}$, we say A is *measurable with respect to* $\mathcal{F}$. We interpret this by saying that $A \in \mathcal{F}$ if we know whether $X \in A$ or not with the partial information in

$\mathcal{F}$. That is, $\mathcal{F}$ is the set of events determined by the partial information. In the above example, suppose the partial information is $X_{[0,t]}$. If $t > 5$, then we know whether $|X_t| < 2$ for all $t \leq 5$, that is, whether $X \in A$. If $t < 5$, the partial information may not decide with certainty whether $X \in A$.

This partial information picture is not a mathematical definition. The definition, which is motivated by the partial information picture, is that an algebra of sets is a family of events that satisfy the following axioms. Set terminology is A^c is the complement of A, so $X \in A$ if and only if $X \notin A^c$. The intersection of events A and B is $X \in A \cap B$ if $X \in A$ and $X \in B$. The union of events is $A \in A \cup B$ if $X \in A$ or $X \in B$ or both. The whole probability space, the space of all possible paths or whatever, is called Ω. The event with no elements is $\emptyset$.

- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$. If we know whether $X \in A$, then we know whether $X \notin A$.
- If A and B are measurable with respect to $\mathcal{F}$, then $A \cup B$ and $A \cap B$ are also measurable with respect to $\mathcal{F}$.
- $\Omega \in \mathcal{F}$ and $\emptyset \in \mathcal{F}$. We know whether $X \in \Omega$ (it is) and whether $X \in \emptyset$ (it isn't).

You can specify an algebra of sets by specifying some of the components or values that define X. We are particularly interested in the case where X is a path and the partial information is a beginning part of the path. The algebra determined by knowing $X_{[0,t]}$ is written $\mathcal{F}_t$. In the example above, $A \in \mathcal{F}_t$ if $t \geq 5$ and $A \notin \mathcal{F}_t$ if $t < 5$. These algebras have the property you gain information with time. That is, if $t' > t$, then $\mathcal{F}_t \subseteq \mathcal{F}_{t'}$. For example $A \in \mathcal{F}_6$, which implies that $A \in \mathcal{F}_7$ because $6 < 7$. An *expanding family of algebras* like this is a *filtration*.

An algebra of events is a $\sigma-$*algebra* if you can take limits of sets within the algebra. You can say this using unions or intersections or both. With unions, we suppose that $A_n \in \mathcal{F}$ for all A. If A is a $\sigma-$algebra, then

$$\cup_{n=1}^{\infty} A_n \in \mathcal{F} \ .$$

The infinite union is defined in a natural way. An outcome X is in the infinite union if it is in any of the A_n. In technical proofs, it is usually necessary for the algebra to be a $\sigma-$algebra. We're not doing proofs here, but you will hear people talk about $\sigma-$algebras and should be aware that it's a technical kind of algebra. If you're not doing proofs, you don't to have to know the technicality.

We say that a function of X is *measurable with respect to* $\mathcal{F}$ is the value of V is determined by the information in $\mathcal{F}$. That is, for any number a, we know whether $V(X) \leq a$.

$$\{X \mid V(X) \leq a\} \in \mathcal{F} \ .$$

It is common to write $V \in \mathcal{F}$ when V is measurable with respect to $\mathcal{F}$. It is clear that you can combine (add, multiply, etc.) measurable functions to get other measurable functions.

Here is the fancy definition of conditional expectation that uses all these ideas. Let $\mathcal{F}$ be a $\sigma-$algebra and V be some function, not necessarily measurable with respect to $\mathcal{F}$. The conditional expectation is a function $\widetilde{V}$ that is measurable with respect to $\mathcal{F}$. We write this as

$$\widetilde{V} = \mathrm{E}[\, V(X) \mid \mathcal{F}] \; .$$

This is defined by either of two properties. One is that if W is any function measurable with respect to $\mathcal{F}$, then

$$\mathrm{E}[\, \widetilde{V}(X)W(X)] = \mathrm{E}[\, V(X)W(X)] \; .$$

The other is that $\widetilde{V}$ is the function measurable with respect to $\mathcal{F}$ that is the best approximation to V in the least squares sense. Exercise 1 explains this in an example.

Some definitions we have used before are often said in the language of conditional expectation. One is the basic value function

$$\mathrm{E}[\, V(X_T) \mid \mathcal{F}_t] = f(X_t, t) \; .$$

The right side involves a variable X_t that is measurable in $\mathcal{F}_t$. The right side involves only X_t because of the Markov property. The *tower property* is the statement that if you take conditional expectation of a conditional expectation, you get the conditional expectation. If $\mathcal{G}$ has more information than $\mathcal{F}$, then

$$\mathrm{E}[\, \mathrm{E}[\, V(X) \mid \mathcal{G}] \mid \mathcal{F}] = \mathrm{E}[\, V(X) \mid \mathcal{F}] \; .$$

In particular, you apply this to the filtration $\mathcal{F}_t$, and $t_2 > t_1$ (more information), you get

$$\mathrm{E}[\, \mathrm{E}[\, V(X) \mid \mathcal{F}_{t_2}] \mid \mathcal{F}_{t_1}] = \mathrm{E}[\, V(X) \mid \mathcal{F}_{t_1}] \; .$$

You can apply this to value functions $V(X_T)$ and get

$$\mathrm{E}[\, \mathrm{E}[\, V(X_T) \mid \mathcal{F}_{t_2}] \mid \mathcal{F}_{t_1}] = \mathrm{E}[\, V(X_T) \mid \mathcal{F}_{t_1}] \; .$$

In terms of value functions, this is

$$\mathrm{E}[\, f(X_{t_2}, t_t) \mid \mathcal{F}_{t_1}] = f(X_{t_1}, t_1) \; .$$

A random process a_t is progressively measurable with respect to the filtration $\mathcal{F}_t$ if a_t is measurable with respect to $\mathcal{F}_t$ for all t. This is the hypothesis you could use in the definition of the Ito integral

$$\int_0^t a_s dW_s \; .$$

A random process Y_t that is progressively measurable is a *martingale* if, whenever $T > t$,

$$Y_t = \mathrm{E}[\, Y_T \mid \mathcal{F}_t] \; .$$

2 Multi-component diffusions

Most dynamic models involve more than one model variable. If X_t is the state of the "system" (whatever is being modeled) at time t, then X_t is likely to have more than one component. We write this as

$$X_t = \begin{pmatrix} X_{1,t} \\ \vdots \\ X_{d,t} \end{pmatrix} .$$

For example, you could descriobe the random motion of a leaf floating on a pond by modeling it's two coordinates as functions of time. In financial markets, you could model the movement of more than one asset price, or more then one currency exchange rate.

The material in this section seems to defy linear ordering. You may have to read it a few times for the different parts to make sense together. If you don't understand something on the first reading, just keep going. It's not hard but it might take some patience. Whenever I see something in vector or matrix notation that I can't understand, I try writing it using indices and sums. The calculations with the multi-component Ito's lemma, such as Exercise 4 may seem complicated, but if you do them slowly it should make sense.

A multi-dimensional diffusion process may be described by giving its infinitesimal mean (drift) and infinitesimal covariance (quadratic variation). In the language of Section 1, we could write these as

$$a(X_t,t)\,dt = \mathrm{E}[\,dX_t \mid \mathcal{F}_t] \ . \tag{4}$$

and

$$\mu(X_t)\,dt = \mathrm{cov}(\,dX_t \mid \mathcal{F}_t) = \mathrm{E}\big[\,dX_t(dX_t)^t \mid \mathcal{F}_t\big] \ . \tag{5}$$

The infinitesimal mean $a(x,t)$ is a $d-$component vector. The infinitesimal covariance $\mu(x,t)$ is a $d \times d$ symmetric positive semi-definite matrix. Recall from the one component case that the second and third expressions of (5) are the same because the difference between them is "tiny".

The formulas with differentials (4) and (5) can be expressed in equivalent "little oh" terms. Let $\Delta t > 0$ be a small increment of time and $\Delta X = X_{t+\Delta t} - X_t$ the corresponding increment of the process. The infinitesimal mean formula (4) is equivalent to

$$a(X_t,t)\Delta t = \mathrm{E}[\,\Delta X \mid \mathcal{F}_t] + o(\Delta t) \ . \tag{6}$$

The infinitesimal covariance formula (5) is equivalent to

$$\mu(X_t)\,\Delta t = \mathrm{cov}(\,\Delta X \mid \mathcal{F}_t) + o(\Delta t) = \mathrm{E}\big[\,\Delta X_t(\Delta X_t)^t \mid \mathcal{F}_t\big] + o(\Delta t) \ . \tag{7}$$

The last two expressions on the right of (7) are not equal but are related by

$$\mathrm{cov}(\,\Delta X \mid \mathcal{F}_t) = \mathrm{E}\big[\,\Delta X_t(\Delta X_t)^t \mid \mathcal{F}_t\big] - \overline{\Delta X}\left(\overline{\Delta X}\right)^t \ , \quad \overline{\Delta X} = \mathrm{E}[\,\Delta X \mid \mathcal{F}_t] \ .$$

From (6), we find that $\overline{\Delta X}\left(\overline{\Delta X}\right)^t = O(\Delta t^2)$. This is the same thing we said when talking about infinitesimal variance of one component processes. A diffusion process model of a random process would be giving $a(x,t)$ and $\mu(x,t)$. This is a stochastic process version of an ordinary differential equation model, which would involve giving just $a(x,t)$.

Diffusion models can be specified using stochastic differential equation systems. The expression looks the same as for a single component SDE, but the objects are different

$$dX_t = a(X_t)\,dt + b(X_t)\,dW_t \; . \tag{8}$$

The drift coefficient $a(x)$ is a $d-$component vector. The noise "coefficient" $b(x)$ is a $d \times m$ matrix. In components, the SDE is

$$dX_{i,t} = a_i(X_t)\,dt + \sum_{j=1}^{m} b_{ij}(X_t)\,dW_{j,t} \; .$$

The number m is the number of *sources of noise*. There must be at least one to make the differential equation stochastic, but there do not have to be as many as there are components of X. Some common models have $m = d$ and others have $m < d$. The latter are called *degenerate diffusions*, but there is nothing wrong with them. The relation between SDE noise coefficient matrix b and infinitesimal covariance matrix μ is found by substituting the SDE (8) into the definition (5). We neglect terms with dt^2 or $dW dt$. We use the fact of matrix and vector algebra, including the fact that transpose reverses the order and matrix/vector multiplication is associative.

$$\begin{aligned}
\mathrm{E}\big[\, dX(dX)^t \mid \mathcal{F}_t \big] &= \mathrm{E}\Big[\, (b(X_t)dW_t)\,(b(X_t)dW_t)^t \Big] \\
&= \mathrm{E}\Big[\, b(X_t)\,dW_t\,(dW_t)^t\,(b(X_t))^t \Big] \\
&= \mathrm{E}\Big[\, b(X_t)\ \Big[dW_t\,(dW_t)^t\Big]\, b(X_t)^t \Big] \\
&= b(X_t)\,\mathrm{E}\Big[\, dW_t\,(dW_t)^t \Big]\, b(X_t)^t \\
&= b(X_t)\,(I_{m\times m}\,dt)\,b^t(X_t) \\
\mu(X_t)\,dt &= b(X_t)b^t(X_t)\,dt
\end{aligned}$$

This shows that the relation between the infinitesimal covariance μ and the noise coefficient b is

$$\mu(x) = b(x)\,b^t(x) \; . \tag{9}$$

If you have a model in terms of $\mu(x)$, you have to find an appropriate $b(x)$, which is somewhat arbitrary, see Exercise 3.

The multi-component Ito's lemma is like the one component Ito's lemma. You have $f(X_t,t)$ and you want df. You expand in a Taylor series until all the terms you have left off are "tiny". Then you replace quadratic expressions like

$dX_{i,t}dXj,t$ by their expected values. That is $\mu_{ij}(X_t)dt$. That's it

$$df(X_t,t) = \sum_{i=1}^{d} \partial_{x_i} dX_i + \partial_t f dt + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \partial_{x_i}\partial_{x_j} f \mu_{ij}(X_t)dt \ . \tag{10}$$

Comments: in the double sum on the right, if $i \neq j$ there are terms (i,j) and (j,i), which are the same. If you group them together, the $\frac{1}{2}$ cancels. There is a $\frac{1}{2}$ only for the diagonal terms, which have $i=j$. If you have an SDE or some other way to describe dX, you can substitute it on the right to get formulas in terms of dW.

One market factor stock process

Let $S_{i,t}$ be the price of asset i at time t, and let d be the number of assets considered. We suppose that each individual asset price is a geometric Brownian motion with some expected rate of return and volatility. We seek a model of correlation between asset prices. One is related to the *one factor* market model. In this model, each asset is driven by its own *idiosyncratic factor* and by a common *market factor*. Stock i has an idiosyncratic noise (a noise that applies only to it), which we call $W_{i,t}$. The noise for the market factor is $W_{0,t}$. This means that there are $d+1$ sources of noise for d components of our diffusion process, which contradicts something above. Please be patient with that. The model is, for $i = 1, \cdots, d$,

$$dS_{i,t} = \mu_i S_{i,t}\, dt + \sigma_i S_{i,t} dW_{i,t} + \beta_i S_{i,t} dW_{0,t} \ . \tag{11}$$

Here is the model written out for $d=3$.

$$d\begin{pmatrix} S_{1,t} \\ S_{2,t} \\ S_{3,t} \end{pmatrix} = \begin{pmatrix} \mu_1 S_{1,t} \\ \mu_2 S_{2,t} \\ \mu_3 S_{3,t} \end{pmatrix} dt + \begin{pmatrix} \sigma_1 S_{1,t} & 0 & 0 & \beta_1 S_{1,t} \\ 0 & \sigma_2 S_{2,t} & 0 & \beta 2 S_{2,t} \\ 0 & 0 & \sigma_{3,t} S_{3,t} & \beta_3 S_{3,t} \end{pmatrix} \begin{pmatrix} dW_{1,t} \\ dW_{2,t} \\ dW_{3,t} \\ dW_{0,t} \end{pmatrix} \ .$$

The infinitesimal mean (drift) is given by

$$a(s) = \begin{pmatrix} \mu_1 s_1 \\ \mu_2 s_2 \\ \mu_3 s_3 \end{pmatrix} \ .$$

The noise coefficient matrix is

$$b(s) = \begin{pmatrix} \sigma_1 s_1 & 0 & 0 & \beta_1 s_1 \\ 0 & \sigma_2 s_2 & 0 & \beta 2 s_2 \\ 0 & 0 & \sigma_{3,t} s_3 & \beta_3 s_3 \end{pmatrix}$$

The infinitesimal covariance is

$$\mu(s) = b(s)b^t(s)$$
$$= \begin{pmatrix} \sigma_1 s_1 & 0 & 0 & \beta_1 s_1 \\ 0 & \sigma_2 s_2 & 0 & \beta 2 s_2 \\ 0 & 0 & \sigma_{3,t} s_3 & \beta_3 s_3 \end{pmatrix} \begin{pmatrix} \sigma_1 s_1 & 0 & 0 \\ 0 & \sigma_2 s_2 0 & \\ 0 & 0 & \sigma_3 s_3 \\ \beta_1 s_1 & \beta_2 s_2 & \beta_3 s_3 \end{pmatrix}$$
$$\mu(s) = \begin{pmatrix} \left(\sigma_1^2 + \beta_1^2\right) s_1^2 & \beta_1 \beta_2 s_1 s_2 & \beta_1 \beta_3 s_1 s_3 \\ \beta_1 \beta_2 s_1 S_2 & \left(\sigma_2^2 + \beta_2^2\right) s_2^2 & \beta_2 \beta_3 s_2 s_3 \\ \beta_1 \beta_3 s_1 s_3 & \beta_2 \beta_3 s_2 s_3 & \left(\sigma_3^2 + \beta_3^2\right) s_3^2 \end{pmatrix} .$$

This matrix is the sum of a diagonal matrix corresponding to independent idiosyncratic factors and a rank one matrix corresponding to the single market factor.

$$\mu(s) = \begin{pmatrix} \sigma_1^2 s_1^2 & 0 & \\ 0 & \sigma_2^2 s_2^2 & \\ 0 & 0 & \sigma_3^2 s_3^2 \end{pmatrix} + \begin{pmatrix} \beta_1 s_1 \\ \beta_2 s_2 \\ \beta_3 s_3 \end{pmatrix} \begin{pmatrix} \beta_1 s_1 & \beta_2 s_2 & \beta_3 s_3 \end{pmatrix} .$$

The last term on the right has the form uu^t, where u is a column vector. Any symmetric rank one matrix has this form.

You can use the multi-dimensional form of Ito's lemma to find a solution formula for the SDE system eqrefofm. Here is a sort-of ansatz approach that is based on our experience with simple geometric Brownian motion. There are other ways to approach this that we will explore below. Suppose the solution has the form

$$S_{i,t} = S_{i,0} e^{\sigma_i W_{i,t} + \beta_i W_{0,t}} M(t) . \tag{12}$$

We apply Ito's lemma with

$$f(w_i, w_0, t) = S_{i,0} e^{\sigma_1 w_i + \beta_i w_0} M(t).$$

This has partial derivatives

$$\partial_{w_i} f = \sigma_i f$$
$$\partial_{w_0} f = \beta_i f$$
$$\partial_t f = S_{i,0} e^{\sigma_1 w_i + \beta_i w_0} \dot{M}(t) = \frac{\dot{M}(t)}{M(t)} f$$
$$\partial_{w_i}^2 f = \sigma_i^2 f$$
$$\partial_{w_0}^2 f = \beta_0^2 f$$
$$\partial_{w_i} \partial_{w_0} f = \sigma_i \beta_i f$$

In this calculation, we leave out the arguments of f and then use $S_{i,t} = f(W_{i,t}, W_{0,t}, t)$. We also use the "Ito rule" formulas $(dW_{i,t})^2 = dt$, $(dW_{0,t})^2 =$

dt, and $(dW_{i,t}dW_{0,t}) = 0$.

$$\begin{aligned}
df(W_{i,t}, W_{0,t}, t) &= \partial_{w_i} f dW_{i,t} + \partial_{w_0} f dW_{0,t} + \partial_t f dt \\
&\quad + \frac{1}{2}\partial^2_{w_i} f (dW_{i,t})^2 + \frac{1}{2}\partial^2_{w_0} f (dW_{0,t})^2 + \partial_{w_i}\partial_{w_0} f (dW_{i,t}dW_{0,t}) \\
&= \sigma_i S_{i,t} dW_{i,t} + \beta_i S_{i,t} dW_{0,t} + \frac{\dot{M}(t)}{M(t)} S_{i,t}\, dt \\
&\quad + \frac{1}{2}\sigma_i^2 S_{i,t} dt + \frac{1}{2}\beta_i S_{i,t} dt
\end{aligned}$$

You can compare this to the SDE (11) and see that the $dW_{i,t}$ and $dW_{i,0}$ terms match. This leaves the dt terms. On the left side are the dt terms from the SDE (11) and on the right are the dt terms from the Ito calculation

$$\mu_i S_{i,t} dt = \frac{\dot{M}(t)}{M(t)} S_{i,t}\, dt + \frac{1}{2}\sigma_i^2 S_{i,t} dt + \frac{1}{2}\beta_i S_{i,t} dt \ .$$

This simplifies to

$$\frac{\dot{M}(t)}{M(t)} = \mu_i - \frac{1}{2}\left(\sigma_i^2 + \beta_i^2\right) \ .$$

We need $M(0) = 1$ so that the ansatz (12) has $S_{i,0} = S_{i,0}$. Therefore,

$$M(t) = e^{\left[\mu_i - \frac{1}{2}\left(\sigma_i^2 + \beta_i^2\right)\right] t} \ .$$

The solution is

$$S_{i,t} = S_{i,0} e^{\sigma_i W_{i,t} + \beta_i W_{0,t} + \left[\mu_i - \frac{1}{2}\left(\sigma_i^2 + \beta_i^2\right)\right] t} \ . \tag{13}$$

A short rate money market model

Let R_t be a market interest rate for deposits (loans) with no default risk and for a very short period. Let M_t be the *money market* account, which is an account that gets the short rate R_t. This evolves M_t according to

$$dM_t = R_t M_t dt \ .$$

Let us assume a linear mean-reverting equilibrium model for the short rate, as we did before

$$dR_t = -\gamma(R_t - \overline{r})dt + \sigma dW_t \ .$$

The parameter γ is the mean reversion rate, $\overline{r}$ is the equilibrium interest rate, and σ determines the amount by which R_t typically differs from $\overline{r}$. This may be written as a two component SDE

$$d\begin{pmatrix} M_t \\ R_r \end{pmatrix} = \begin{pmatrix} R_t M_t \\ -\gamma(R_t - \overline{r}) \end{pmatrix} dt + \begin{pmatrix} 0 \\ \sigma \end{pmatrix} dW_t \ .$$

This model has $d = 2$ components and $m = 1$ sources of noise. The infinitesimal covariance is

$$\mu = bb^t = \begin{pmatrix} 0 \\ \sigma \end{pmatrix} \begin{pmatrix} 0 & \sigma \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix} \ .$$

3 Exercises

These exercises are for study and learning only. Do them as you have time. Do not hand them in. They will not be graded.

1. Suppose X is a one component random variable with probability density $u(x)$. Show that the conditional expectation of $V(X)$ is the number that best approximates $V(X)$ in the least squares sense. That is if $a = \mathrm{E}[V(X)]$, and $b \neq a$, then
$$\mathrm{E}\big[\,[(V(X) - a)^2\big] < \mathrm{E}\big[\,[(V(X) - b)^2\big] \ .$$
Hint. Put $b = a + (b - a)$ on the right and expand the square. Now suppose $X = (X_1, X_2)$ and let $\mathcal{F}$ be the $\sigma-$algebra of events that "know X_1". Let $u(x_1, x_2)$ be the PDF of (X_1, X_2). Find a formula for
$$\widetilde{V}(X_1) = \mathrm{E}[\,V(X_1, X_2) \mid \mathcal{F}]$$
Show that if W is measurable with respect to $\mathcal{F}$ then
$$\mathrm{E}\Big[\,(V(X_1, X_2) - \widetilde{V}(X_1))^2\Big] < \mathrm{E}\big[\,(V(X_1, X_2) - W(X_1))^2\big] \ .$$
This shows that the conditional expectation is the function measurable in $\mathcal{F}$ that best approximates V in the least squares sense.

2. Let X_t be one dimensional Brownian motion and $V(x) = x^4$. Calculate the value function $f(x, t)$. Show by explicit calculation that if $t_2 > t_1$, then
$$\mathrm{E}[\,f(X_{t_2}, t_2) \mid \mathcal{F}_{t_1}] = f(X_{t_1}, t_1) \ .$$

3. There can be different ways to "explain" correlations between multi-variate Gaussians, or random variables with other distributions. Suppose $X = (X_1, X_2)$ is a two component random variable so that $X_1 = \sigma_1 Z_1$ and $X_2 = \rho X_1 + \sigma_2 Z_2$. Here, Z_1 and Z_2 are independent standard normals. These formulas explain the correlation between X_1 and X_2 by imagining that X_1 "drives" X_2 with a coefficient ρ. Calculate the covariance matrix
$$C = \mathrm{cov}(X) = \mathrm{E}\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \begin{pmatrix} X_1 & X_2 \end{pmatrix}\right] \ .$$
Find a different explanation for X in which X_2 is defined first in terms of a standard normal $X_2 = \widetilde{\sigma}_2 Z_2$ and then X_1 is driven by X_2 with some extra noise added: $X_1 = \widetilde{\rho} X_2 + \widetilde{\sigma}_1 Z_1$. Find an upper-triangular matrix b and a lower triangular matrix $\widetilde{b}$ so that
$$C = bb^t = \widetilde{b}\widetilde{b}^t \ .$$
Conclude that the noise coefficient b in the SDE (8) is not determined by the infinitesimal mean and covariance of the process X_t.

4. The solution formula in the one factor multi-asset price model (13) has the form
$$S_{i,t} = \widetilde{S}_{i,t} U_t \ ,$$
where the idiosyncratic part is a geometric Brownian motion involving the idiosyncratic noise only
$$d\widetilde{S}_{i,t} = \mu_i \widetilde{S}_{i,t} dt + \sigma_i \widetilde{S}_{i,t} dW_{i,t} \ , \quad \widetilde{S}_{i,0} = S_{i,0} \ ,$$
and the market part is a geometric Brownian motion with just the market factor noise
$$dU_{i,t} = \beta_i U_{i,t} dW_{0,t} \ , \quad U_{i,0} = 1 \ .$$
Verify this product form directly using the multi-component Ito's lemma.