PROBABILITY 2018
HANDOUT 4: MARTINGALES, CONDITIONAL INDEPENDENCE

GIOVANNI PISTONE

Contents

1. Martingales

1. A finite or infinite sequence $X_0, X_1, \dots$ of integrable real random variables of the probability space $(\Omega, \mathcal{F}, \mu)$ is a *martingale* if all $k \geqslant 1$ in the index set it holds

$$\mathrm{E}\left(X_k | X_i \colon i < k\right) = X_{k-1}$$

that is,

$$\mathbb{E}\left(X_k g(X_0, \dots, X_{k-1})\right) = \mathbb{E}\left(X_{k-1} g(X_0, \dots, X_{k-1})\right)$$

for all bounded measurable g. Notice that the property of being a martingale refers to a condition on the conditional distribution of the variables in the sequence with respect to the past namely, if $\mu_{X_k|X_0,\dots,X_{k-1}}$ is the conditional distribution of X_k given $X_0, \dots, X_{k-1}$, then the martingale condition is

$$\int y \ \mu_{X_k|X_0,\dots,X_{k-1}}(dy|X_0, \dots, X_{k-1}) = X_{k-1} \quad \mu\text{-a.s.}$$

The theory of martingales is fully developed in [1, Part B]. Here we discuss only a few basic fact in form of exercises.

Exercise 1. (1) Let X, Y be Bernoulli variables with $p_{X,Y}(x, y) = \mathrm{P}\left(X = x, Y = y\right)$, $x, = 0, 1$. Write the condition on the joint probability function $p_{X,Y}$ equivalent to $[X, Y]$ being a martingale.
(2) Let $[X \ Y]^T \sim \mathrm{N}_2\left(\mu, \Sigma\right)$. Write the condition on μ and Σ equivalent to $[X, Y]$ being a martingale.
(3) Let be given a sequence of measurable functions $X_0, \dots, X_n$ of the measurable space $(\Omega, \mathcal{F})$. The set of all probabilities such that the given sequence is a martingale is a convex set.
(4) Let $\mathcal{F}_k$, $k = 0, \dots, n$, be an increasing sequence of sub-σ-algebras of $\mathcal{F}$. Such a sequence is called a *filtration*. The sequence X_k, $k = 0, \dots n$, is *adapted* to the given filtration if each X_k is $\mathcal{F}_k$-measurable. Each sequence is adapted to the *natural filtration* $\mathcal{F}_k = \sigma X_0, \dots, X_k$, $k = 0, \dots n$. Show that the sequence is a martingale if

$$\mathrm{E}\left(X_k | \mathcal{F}_{k-1}\right) = X_{k-1} \quad k \geqslant 1 \ .$$

Date: April 11, 2018 PROVISIONAL VERSION.

(5) Assume the sequence of real integrable random variables $[X_0, \ldots, X_n]$ is adapted to the filtration $(\mathcal{F}_k)_{k=0}^n$. Then, it is a martingale if, and only if, $\mathrm{E}\left(X_k - X_{k-1} | \mathcal{F}_{k-1}\right) = 0$.

(6) Let $Z_1, \ldots$ be a sequence of independent integrable real random variables of the probability space $(\Omega, \mathcal{F}, \mu)$ and assume $\mathbb{E}\left(Z_k\right) = 0$ for all k. The sequence $X_k = \sum_{i \leqslant k}$, $k = 0, 1, \ldots$, is the *symmetric random walk*. Show that $X_0, X_i, \ldots$ is a martingale. Let $\phi \colon \mathbb{R} \to \mathbb{R}$ be a bounded measurable function. Under which condition the sequence $\phi(X_0), \phi(X_1), \ldots$ is a martingale?

(7) Let $Z_1, \ldots, Z_n$ be a sequence of independent integrable real random variables of the probability space $(\Omega, \mathcal{F}, \mu)$. Let be given a initial value $x_0 \in \mathbb{R}$ and sequence of bounded measurable functions $s_k \colon \mathbb{R}^{k-1} \to \mathbb{R}$. Let us call such a sequence a *strategy*. Define $X_0 = x_0$ and $X_k = X_{k-1} + Z_k s_k(X_0, \ldots, X_{k-1})$. Such a sequence is called the *total gain*. Show that it is a martingale for any strategy if, and only if, $\mathbb{E}\left(Z_k\right) = 0$, $k \geqslant 1$. Show that in fact it is enough to assume that $\mathrm{E}\left(Z_k | \mathcal{F}_{k-1}\right) = 0$. *No non-anticipating strategy* can transform a martingale into something else.

(8) Consider a martingale X_k, $k = 0, \ldots n$ for the filtration $\mathcal{F}_k$, $k = 0, \ldots, n$. A *stopping time* or *optional time* is a random variable T with values in $\{0, \ldots, n\} \cup \infty$ such that $\{T = k\} \in \mathcal{F}_k$, $k = 0, \ldots, n$. Give an example of stopping time. The *stopped process* is defined by $X_k^T = X_{T \wedge k}$. Show that the stopped process is a martingale for the same filtration as the original one. The stopped process is a martingale based on a strategy.

(9) A Gaussian vector $X = [X_0 \cdots X_n]$ is a martingale if, and only if, the increments $X_k - X_{k-1}$, $k \geqslant 1$, are independent. Compute the distribution of X. Which are the free parameters in the distribution of X? [Hint: Consider the increments $Z_k = X_k - X_{k-1}$, $k = 1, \ldots, n$. If $(\mathcal{F}_k)_{k=0}^n$ be the natural filtration. We want $\mathrm{E}\left(Z_k | \mathcal{F}_{k-1}\right) = 0$, $k \geqslant 1$. The filtration generated by $X_0, Z_1, \ldots$ is equal to the natural filtration of $[X_k]_{k=1}^n$. It follows that the sequence $x_0, Z_1, \ldots$ is independent.]

2. Conditional independence

Conditional independence is a key property in Statistics e.g. Graphical Models, in Stochastic Processes e.g., Markov processes, in Random Fields, in Machine Learning.

2 (Conditional independence)**.**

(1) The non-null events A, B, C are such that A and C are independent given B, $A \perp\!\!\!\perp C |\, B$, if each one of the following equivalent conditions are satisfied:

$$\mathrm{P}\left(A \cap C | B\right) = \mathrm{P}\left(A | B\right) \mathrm{P}\left(C | B\right)$$
$$\mathrm{P}\left(A | B \cap C\right) = \mathrm{P}\left(A | B\right)$$
$$\mathrm{P}\left(A \cap B \cap C\right) \mathrm{P}\left(B\right) = \mathrm{P}\left(A \cap B\right) \mathrm{P}\left(B \cap C\right)$$

Notice that the last condition is meaningful even if some of the events has is a null event.

(2) Random variables Y_1, Y_3 are conditionally independent given the random variable Y_2, $Y_1 \perp\!\!\!\perp Y_3 |\, Y_2$ if each one of the following equivalent conditions are satisfied. If f_i, $i = 1, \ldots, 3$, are bounded,

$$\mathrm{E}\left(f_1(Y_1) f_3(Y_3) | Y_2\right) = \mathrm{E}\left(f_1(Y_1) | Y_2\right) \mathrm{E}\left(f_3(Y_3) | Y_2\right)$$
$$\mathrm{E}\left(f_1(Y_1) | Y_2, Y_3\right) = \mathrm{E}\left(f_1(Y_1) | Y_2\right)$$

(3) Let $\mu_{(Y_1,Y_3)|Y_2}$ be the conditional distribution of (Y_1, Y_3) given Y_3. Then, $Y_1 \perp\!\!\!\perp Y_3 |\, Y_2$ if, and only if, $\mu_{(Y_1,Y_3)|Y_2} = \mu_{Y_1|Y_2} \otimes \mu_{Y_3|Y_2}$.

(4) A stochastic process $Y_1, \dots, Y_N$ is a *Markov Process* if

$$(Y_1, \dots, Y_k) \perp\!\!\!\perp (Y_k, \dots, Y_N) |\, Y_k \ , \quad k = 2, \dots, N-1 \ .$$

Exercise 2. Prove the equivalence of the statements for conditional indendence of events.

Exercise 3. Prove the equivalence of the two statement for conditional independence of random variables.

Exercise 4. Let be given a Gaussian white noise $Z_1, \dots, Z_n$ and a further independent gaussian random variable X_0. For each real α define $X_k = \alpha X_{k-1} + Z_k$, $k \geqslant 1$. Show that it is a Markov process.

Proposition 1. *Let be given*

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} \sim \mathrm{N}_{n_1+n_2+n_3} \left(\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} \right)$$

We have $Y_1 \perp\!\!\!\perp Y_3 |\, Y_2$ if, and only if, $\Sigma_{13} = \Sigma_{12}\Sigma_{22}^{+}\Sigma_{23}$. In such a case,

$$\begin{bmatrix} Y_1 \\ Y_3 \end{bmatrix} \Bigg| (Y_2 = y_2) \sim \mathrm{N}_{n_1+n_3} \left(\begin{bmatrix} b_1 \\ b_3 \end{bmatrix} + \begin{bmatrix} \Sigma_{12} \\ \Sigma_{32} \end{bmatrix} \Sigma_{22}^{+}(y_2 - b_2), \begin{bmatrix} \Sigma_{1|2} & 0 \\ 0 & \Sigma_{3|2} \end{bmatrix} \right)$$

and

$$Y_1 | (Y_2 = y_2, Y_3 = y_3) = Y_1 | (Y_2 = y_2) \sim \mathrm{N}_{n_1} \left(b_1 + \Sigma_{1,2}\Sigma_{22}^{+}(y_2 - b_2), \Sigma_{1|2} \right)$$

Proof. Let us apply the conditioning formula to the partitioned Gaussian vector

$$\begin{bmatrix} Y_1 \\ Y_3 \\ \hline Y_2 \end{bmatrix} \sim \mathrm{N}_{(n_1+n_3)+n_2} \left(\begin{bmatrix} b_1 \\ b_3 \\ \hline b_2 \end{bmatrix}, \left[\begin{array}{cc|c} \Sigma_{11} & \Sigma_{13} & \Sigma_{12} \\ \Sigma_{31} & \Sigma_{33} & \Sigma_{32} \\ \hline \Sigma_{21} & \Sigma_{23} & \Sigma_{22} \end{array}\right] \right) .$$

Let us compute the matrix

$$L_{(13)2} = \Sigma_{(13)2}\Sigma_{22}^{+} = \begin{bmatrix} \Sigma_{12} \\ \Sigma_{32} \end{bmatrix} \Sigma_{22}^{+} = \begin{bmatrix} \Sigma_{12}\Sigma_{22}^{+} \\ \Sigma_{32}\Sigma_{22}^{+} \end{bmatrix}$$

and the conditional variance

$$\begin{aligned} \Sigma_{(13)|2} = \Sigma_{(13)(13)} - L_{(13)2}\Sigma_{2(13)} = \\ \begin{bmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{bmatrix} - \begin{bmatrix} \Sigma_{12}\Sigma_{22}^{+} \\ \Sigma_{32}\Sigma_{22}^{+} \end{bmatrix} \begin{bmatrix} \Sigma_{21} & \Sigma_{23} \end{bmatrix} = \\ \begin{bmatrix} \Sigma_{11} & \Sigma_{13} \\ \Sigma_{31} & \Sigma_{33} \end{bmatrix} - \begin{bmatrix} \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} & \Sigma_{12}\Sigma_{22}^{+}\Sigma_{23} \\ \Sigma_{32}\Sigma_{22}^{+}\Sigma_{21} & \Sigma_{32}\Sigma_{22}^{+}\Sigma_{23} \end{bmatrix} = \\ \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} & \Sigma_{13} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{23} \\ \Sigma_{31} - \Sigma_{32}\Sigma_{22}^{+}\Sigma_{21} & \Sigma_{33} - \Sigma_{32}\Sigma_{22}^{+}\Sigma_{23} \end{bmatrix} \end{aligned}$$

Then $Y_1 \perp\!\!\!\perp Y_3 |\, Y_2$ if, and only if, the conditional variance is block-diagonal, $\Sigma_{13} = \Sigma_{12}\Sigma_{22}^{+}\Sigma_{23}$.

Consider now the partition

$$\begin{bmatrix} Y_1 \\ \hline Y_2 \\ Y_3 \end{bmatrix} \sim \mathrm{N}_{n_1+(n_2+n_3)} \left(\begin{bmatrix} b_1 \\ \hline b_2 \\ b_3 \end{bmatrix}, \left[\begin{array}{c|cc} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \hline \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{array}\right] \right) .$$

We have

$$L_{1(23)} = \Sigma_{1(23)}\Sigma_{(23)(23)}^{+} = \begin{bmatrix}\Sigma_{12} & \Sigma_{13}\end{bmatrix}\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix}^{+}$$

and

$$\Sigma_{1|(23)} = \Sigma_{11} - L_{1(23)}\Sigma_{(23)1} = \Sigma_{11} - \begin{bmatrix}\Sigma_{12} & \Sigma_{13}\end{bmatrix}\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix}^{+}\begin{bmatrix}\Sigma_{21}\\ \Sigma_{31}\end{bmatrix} .$$

The conditional distribution is

$$Y_1|(Y_2 = y_2, Y_3 = y_3) \sim$$
$$\mathrm{N}_{n_1}\left(b_1 + L_{1(23)}\begin{bmatrix}y_2 - b_2\\ y_3 - b_3\end{bmatrix}, \Sigma_{11} - \begin{bmatrix}\Sigma_{12} & \Sigma_{13}\end{bmatrix}\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix}^{+}\begin{bmatrix}\Sigma_{21}\\ \Sigma_{31}\end{bmatrix}\right)$$

We can write the conditional independence condition as

$$\begin{bmatrix}\Sigma_{12} & \Sigma_{13}\end{bmatrix}\begin{bmatrix}I & -\Sigma_{22}^{+}\Sigma_{23}\\ 0 & I\end{bmatrix} = \begin{bmatrix}\Sigma_{12} & -\Sigma_{12}\Sigma_{22}^{+}\Sigma_{23} + \Sigma_{13}\\ 0 & I\end{bmatrix} = \begin{bmatrix}\Sigma_{12} & 0\end{bmatrix}$$

This computation points to the Schur complement lemma. Check first that

$$\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix} = \begin{bmatrix}I & 0\\ \Sigma_{32}\Sigma_{22}^{+} & I\end{bmatrix}\begin{bmatrix}\Sigma_{22} & 0\\ 0 & \Sigma_{3|2}\end{bmatrix}\begin{bmatrix}I & \Sigma_{22}^{+}\Sigma_{23}\\ 0 & I\ .\end{bmatrix}$$

Then check that

$$\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix}^{+} = \begin{bmatrix}I & -\Sigma_{22}^{+}\Sigma_{23}\\ & I\end{bmatrix}\begin{bmatrix}\Sigma_{22}^{+} & 0\\ 0 & \Sigma_{3|2}^{+}\end{bmatrix}\begin{bmatrix}I & \\ -\Sigma_{32}\Sigma_{22}^{+} & I\ .\end{bmatrix}$$

It follows that

$$L_{1(23)} = \begin{bmatrix}\Sigma_{12} & \Sigma_{13}\end{bmatrix}\begin{bmatrix}\Sigma_{22} & \Sigma_{23}\\ \Sigma_{32} & \Sigma_{33}\end{bmatrix}^{+} =$$
$$\begin{bmatrix}\Sigma_{12} & \Sigma_{13}\end{bmatrix}\begin{bmatrix}I & -\Sigma_{22}^{+}\Sigma_{23}\\ 0 & I\end{bmatrix}\begin{bmatrix}\Sigma_{22}^{+} & 0\\ 0 & \Sigma_{3|2}^{+}\end{bmatrix}\begin{bmatrix}I & 0\\ -\Sigma_{32}\Sigma_{22}^{+} & I\end{bmatrix} =$$
$$\begin{bmatrix}\Sigma_{12} & 0\end{bmatrix}\begin{bmatrix}\Sigma_{22}^{+} & 0\\ 0 & \Sigma_{3|2}^{+}\end{bmatrix}\begin{bmatrix}I & 0\\ -\Sigma_{32}\Sigma_{22}^{+} & I\end{bmatrix} = \begin{bmatrix}\Sigma_{12}\Sigma_{22}^{+} & 0\end{bmatrix}\begin{bmatrix}I & 0\\ -\Sigma_{32}\Sigma_{22}^{+} & I\end{bmatrix} =$$
$$\begin{bmatrix}\Sigma_{12}\Sigma_{22}^{+} & 0\end{bmatrix} ,$$

so that,

$$\Sigma_{1|(23)} = \Sigma_{11} - \begin{bmatrix}\Sigma_{12}\Sigma_{22}^{+} & 0\end{bmatrix}\begin{bmatrix}\Sigma_{21}\\ \Sigma_{31}\end{bmatrix} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{+}\Sigma_{21} = \Sigma_{1|2} .$$

□

Exercise 5. Let $Y = (Y_1, Y_2, Y_3, Y_4)$ be a Gaussian vector with zero mean and such that each component is standard. Write the conditions imposed on the elements of the covariance matrix by the Markov property.

Exercise 6 (Martingale problem). (1) Let $Y_0, Y_1, \dots, Y_N$ be a Markov process, each random variable having values in the measurable space $(S, \mathcal{S})$. Given any bounded measurable $\phi\colon S \to \mathbb{R}$ define the new process $X_0^{\phi} = \phi(Y_0)$,

$$X_t^{\phi} = X_{t-1}^{\phi} + \phi(Y_t) - \mathrm{E}\left(\phi(Y_t)|Y_0, \dots, Y_{t-1}\right) , \quad t \geqslant 1 .$$

Then, X_t^{ϕ} is $(Y_0, \dots, Y_t)$-measurable and has the martingale property

$$\mathrm{E}\left(X_t^{\phi}\middle|Y_0, \dots, Y_{t-1}\right) = X_{t-1}^{\phi} .$$

(2) Because the Markov property is a property of conditional independence, we know that

$$\mathrm{E}\left(\phi(Y_t)|Y_0,\ldots,Y_{t-1}\right) = \mathrm{E}\left(\phi(Y_t)|Y_{t-1}\right) = \int \phi(y)\ \mu_{Y_t|Y_{t-1}}(dy|Y_{t-1})\ .$$

The family of operators $\phi \mapsto A_t\phi$, $t = 1,\ldots,N$, defined by $A\phi(x) = \phi(x) - \int \phi(y)\ \mu_{Y_t|Y_{t-1}}(dy|x)$ is called the generator of the Markov process. We have

$$X_t^\phi - X_{t-1}^\phi = A\phi(X_{t-1}^\phi)$$

References

[1] David Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.

Collegio Carlo Alberto
E-mail address: giovanni.pistone@carloalberto.org