RCP and Martingales

12.1 Regular Conditional Probability (RCP)

Given two measure spaces $(\Omega, \mathcal{F})$ and $(S, \mathscr{S})$, a Markov kernel $Q(\omega, A) : \Omega \times \mathscr{S} \to [0,1]$ is a function such that

1. $Q(\cdot, A)$ is $\mathcal{F}$ measurable for fixed $A \in \mathscr{S}$,
2. $Q(\omega, \cdot)$ is a probability measure on $(S, \mathscr{S})$.

For a measurable function $X : (\Omega, \mathcal{F}) \to (S, \mathscr{S})$ and a sub σ-field $\mathcal{G} \subseteq \mathcal{F}$, Regular Conditional Probability (RCP) is a Markov kernel $Q(\omega, A)$ on $\Omega \times \mathscr{S}$ which is a version of $\mathbb{P}(X \in A \mid \mathcal{G}), A \in \mathscr{S}$, *i.e.*,

$$\mathbb{P}(Q(\cdot, A) = \mathbb{P}(X \in A \mid \mathcal{G})) = 1 \text{ for all } A \in \mathscr{S}.$$

Theorem 12.1. *RCP exists for real-valued r.v. X.*

Proof. For $r \in \mathbb{Q}$, $\mathbb{P}(X \leqslant r \mid \mathcal{G})$ is defined a.s. on $\omega \in A_r, \mathbb{P}(A_r) = 1$. Therefore, $A = \cap_{r \in \mathbb{Q}} A_r$ has $\mathbb{P}(A) = 1$ and $F(\omega, r) := \mathbb{P}(X \leqslant r \mid \mathcal{G})$ exists and is monotone in r for all $r \in \mathbb{Q}, \omega \in A$. For $\omega \in A$, define $F(\omega, x) = \inf_{r > x, r \in \mathbb{Q}} F(\omega, r), x \in \mathbb{R}$. Define $Q(w, C)$ as follows:

$$Q(\omega, C) = \begin{cases} \int_C dF(\omega, x) & \text{if } \omega \in A \\ \mathbb{1}_{\{0 \in C\}} & \text{if } \omega \notin A. \end{cases}$$

Then $Q(\cdot, A)$ is a RCP for $\mathbb{P}(X \in A \mid \mathcal{G}), A \in \mathcal{B}$. ■

Lemma 12.2. *$X : (\Omega, \mathcal{F}) \to (S, \mathscr{S}), \mathcal{G} \subseteq \mathcal{F}$. RCP exists if $(S, \mathscr{S})$ is Borel isomorphic to $(\mathbb{R}, \mathcal{B})$. ($\exists$ bimeasurable bijection $(S, \mathscr{S}) \to (\mathbb{R}, \mathcal{B})$).*

Example 12.3. *Let $X \perp \mathcal{G}$, $X, Y : (\sigma, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{R})$ and Y is $\mathcal{G}$-measurable, $\mathcal{G} \subseteq \mathcal{F}$. Find RCP for $X + Y | \mathcal{G}$. We want to define $\mathbb{P}(X + Y \in A | \mathcal{G})$ for $\omega \in \Omega$ and $A \in \mathcal{R}$. Since Y is $\mathcal{G}$-measurable, given $\mathcal{G}$, Y can be think of as a constant. Let $\mu_X(A) = \mathbb{P}(X \in A)$, then RCP for $X + Y | \mathcal{G}$ is $\mu_X(A - Y(\omega))$*

Proof.

1. $\forall A \in \mathcal{R}$, $\mu_X(A - Y(\omega))$ is a function of Y, so it is $\mathcal{G}$-measurable.
2. $\forall \omega \in \Omega$, $\mu_X(A - Y(\omega))$ is a probability measure.
3. $\forall Z$ bounded and $\mathcal{G}$-measurable, want to show that $\mathbb{E}(Z \mu_X(A - Y)) = \mathbb{E}(Z \mathbb{1}_{X+Y \in A})$. We have

$$\mathbb{E}(Z \mu_X(A - Y(\omega))) = \mathbb{E}(Z \int \mathbb{1}_{x + Y(\omega) \in A} \mathbb{P}_X(dx)) = \int \mathbb{E}(Z \mathbb{1}_{x + Y(\omega)}) \mathbb{P}(dx) = \mathbb{E}(Z \mathbb{1}_{X+Y \in A})$$

where the last equality follows by Fubini. ■

Example 12.4. *Suppose (X,Y) has the joint density $f(x,y)dxdy$ and $\mathcal{G}=\sigma(Y)$, then $X|Y = X|\sigma(Y)$ has RCP defined as the following. Marginal density of Y is $f_Y(y)=\int_{\mathbb{R}} f(x,y)dx$. Conditional density $g_Y(x)=\frac{f(x,y)}{f_Y(y)}$. We claim that RCP of $X \mid Y$ is*

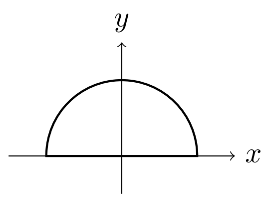
$$\mathbb{P}(X \in A \mid Y) = \int_A \frac{f(x,y)}{f_Y(y)} dx \bigg|_{y=Y(\omega)} \quad \forall \omega \in \Omega \text{ and } A \in \mathcal{B}.$$

Proof. We have $\int_A \frac{f(x,y)}{f_Y(y)}dx\big|_{y=Y(\omega)}$ satisfies

1. Y-measurable (for fixed A).
2. a probability measure (for fixed ω).
3. for all $Z, \sigma(Y)$-measurable and bounded we have, $\mathbb{E}(Z\mathbb{1}_{X\in A}) = \mathbb{E}(Z\int_A \frac{f(x,y)}{f_Y(y)}dx|_{y=Y(\omega)})$. We use the fact that, any Z which is $\sigma(Y)$ measurable must be $\eta(Y)$ for some measurable η. Thus we have

$$\mathbb{E}\left(\int_A \frac{f(x,y)}{f_Y(y)}dx\bigg|_{y=Y}\eta(Y)\right) = \int\left[\int_A \frac{f(x,y)}{f_Y(y)}\eta(y)dx\right] f_Y(y)dy = \mathbb{E}\left(\eta(Y)\mathbb{1}_{X\in A}\right). \qquad \blacksquare$$

Note: Suppose $\mathcal{G}=\sigma(Y)$ and the RCP of $X \mid Y$ is $Q(\omega, A)=\mathbb{P}(X\in A \mid \sigma(Y))$, then this RCP is a function of Y, it must be of the form $\hat{Q}(Y(\omega), A)$. Since RCP is a measurable function of $\mathcal{G}$, and thus it is a function of Y. We will use the notation, $\mathbb{P}(X \in A \mid Y=y)$ for $\hat{Q}(y,A)$.



Example 12.5. *Suppose that a point with Cartesian coordinate (X,Y) is uniformly distributed on the half disk $\{(x,y): y \geqslant 0, x^2+y^2 \leqslant 1\}$, then $\mathbb{P}(Y \leqslant 1/2|X=0)=1/2$. In fact, the RCP of $Y \mid X$ is given by*

$$\mathbb{P}(Y\in A \mid X) = \frac{|A \cap [0,\sqrt{1-x^2}]|}{\sqrt{1-x^2}}\Big|_{x=X}$$

and then evaluate this at $x=0, A=[0,1/2]$.

If $\mathbb{P}(Y\in A \mid X)=\hat{Q}(X(\omega),A)$, then

$$\mathbb{P}(Y\in A \mid X\in(x\pm\varepsilon)) = \frac{\mathbb{P}(Y\in A, X\in(x\pm\varepsilon))}{\mathbb{P}(X\in(x\pm\varepsilon))} = \frac{\mathbb{E}(\hat{Q}(X,A)\mathbb{1}_{X\in(x\pm\varepsilon)})}{\mathbb{P}(X\in(x\pm\varepsilon))}.$$

Thus if RCP exists and x is in the support of X, $\lim_{\varepsilon\to 0}\mathbb{P}(Y\in A \mid X\in(x\pm\varepsilon)) = \hat{Q}(x,A)$.

Note: Rigorously speaking, for set A and set B, $\mathbb{P}(A \mid B)$ is not defined. We only define a conditional probability given a σ-field. Intuitively $\mathbb{P}(A \mid B) \equiv \mathbb{P}(A \mid \sigma(B)=\{\phi, B, B^c, \Omega\})$

Example 12.6. *Think of the previous example in a polar coordinate. Let a point with polar coordinate (R,Θ) be uniformly distributed on a half disk, then $\mathbb{P}(R\leqslant 1/2 \mid \Theta=\pi/2)=(1/2)^2=1/4$. The existence of RCP of $R \mid \Theta$ implies that $\lim_{\varepsilon\to 0}\mathbb{P}(R\leqslant 1/2 \mid \Theta\in(\pi/2\pm\varepsilon))$ exists and equals $\mathbb{P}(R\leqslant 1/2 \mid \Theta=\pi/2)$.*

12.2 Stopping time

Definition 12.7. *A (discrete) filtration is a sequence of increasing σ-fields,*

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots .$$

e.g., Let $X_1, X_2, \ldots$ be i.i.d r.v.s. Then $\mathcal{F}_n = \sigma(X_1, \ldots, X_n), n \geqslant 0$ is a filtration.

Definition 12.8. *A sequence of events $(A_n)_{n\geqslant 0}$ is adapted to the filtration $(\mathcal{F}_n)_{n\geqslant 0}$ if,*

$$A_n \in \mathcal{F}_n, \quad \forall\, n \geqslant 0.$$

A sequence of r.v.s $(X_n)_{n\geqslant 0}$ is adapted to the filtration $(\mathcal{F}_n)_{n\geqslant 0}$ if,

$$\sigma(X_n) \subseteq \mathcal{F}_n, \quad \forall\, n \geqslant 0.$$

Definition 12.9. *A sequence of events $(A_n)_{n\geqslant 0}$ is predictable w.r.t. the filtration $(\mathcal{F}_n)_{n\geqslant 0}$ if,*

$$A_{n+1} \in \mathcal{F}_n, \quad \forall\, n \geqslant 0.$$

A sequence of r.v.s $(X_n)_{n\geqslant 1}$ is predictable w.r.t. to the filtration $(\mathcal{F}_n)_{n\geqslant 1}$ if,

$$\sigma(X_{n+1}) \subseteq \mathcal{F}_n, \quad \forall\, n \geqslant 0.$$

Definition 12.10. *Stopping time w.r.t. a filtration $(\mathcal{F}_n)_{n\geqslant 0}$ is a r.v. $T : \Omega \to \{0, 1, 2, \ldots\}$ s.t.,*

$$\{T = n\} \in \mathcal{F}_n \quad \forall\, n \geqslant 0.$$

Example 12.11. *Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.s and $S_n = X_1 + X_2 + \cdots + X_n, n \geqslant 0$. Then*

$$T_A = \inf\{n |\ S_n \in A\}$$

is a stopping time (hitting time) with respect to the standard filtration $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$.

Proof. We have $\{T = n\} = \{S_n \in A;\ S_i \notin A\ , i = 1, \ldots, n-1\} \in \sigma(X_1, \ldots, X_n) = \mathcal{F}_n$. ∎

Example 12.12. *Let $X_1, \ldots, X_N$ be i.i.d. r.v.s. Then*

$$T = \min\{i \mid X_i = \max(X_1, \ldots, X_N)\}$$

is not a stopping time w.r.t. $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$.

Proof. We have $\{T = n\} = \{X_1, X_2, \ldots, X_{n-1} < X_n, X_{n+1}, \ldots, X_N \leqslant X_n\} \notin \sigma(X_1, \ldots, X_n)$. ∎

Example 12.13. *Consider Example 12.11 with $\mathcal{F}_n := \sigma(X_1, \ldots, X_{n+1})$. Then $\{T = n\}$ is predictable.*

Lemma 12.14. *T is stopping time w.r.t filtration $\mathcal{F}_n$ iff*

$$\{T \leqslant n\} \in F_n, \quad or \quad \{T > n\} \in \mathcal{F}_n, \quad \forall\, n.$$

Lemma 12.15. *If T and S are two stopping time w.r.t the same filtration $(\mathcal{F}_n)$, then $S+T$, $S \vee T$ and $S \wedge T$ are stopping times.*

Proof. For $S + T$, we have $\{S + T = n\} = \cup_{k=0}^{n}(\underbrace{\{S = k\}}_{\in \mathcal{F}_k \subset \mathcal{F}_n} \cap \underbrace{\{T = n - k\}}_{\in \mathcal{F}_{n-k} \subset \mathcal{F}_n}) \in \mathcal{F}_n$. ∎

Example 12.16. *Consider the setup in Example 12.11. Then,*

$$T_A \wedge T_B = T_{A\cup B} \text{ and } T_A \vee T_B = T_{A\cap B}$$

12.3 Martingale

After studying random variables and random vectors, a natural next step is a countably infinite vector of random variables $(X_n)_{n\geqslant 0}$. We can think of the sequence as a time indexed sequence where the time is in $\mathbb{Z}_+ = \{0,1,2,\ldots\}$. As we measure information by σ-fields, we can define $\mathcal{F}_n$ as the amount of information available at time $n \geqslant 0$ and naturally we have $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all n, which is the definition of a filtration. Also, X_n should be $\mathcal{F}_n$ measurable for all n. By abstract result (Ionescu-Tulcea Extension theorem) it follows that, if we know the conditional distribution of X_{n+1} given $\sigma(X_i, 0 \leqslant i \leqslant n)$ for all n, we can reconstruct the whole infinite process. There are mainly two direction one can take: a) assume that the dependence on the past is weak, say the conditional distribution of X_{n+1} given $\sigma(X_i, 0 \leqslant i \leqslant n)$ depends only on X_n (or n, X_n) for all n, which gives rise to Markov Chains (or time-inhomogeneous Markov chain), b) instead of compromising on the dependence on the past, we assume the conditional expectation (or some other statistic) of X_{n+1} given $\sigma(X_i, 0 \leqslant i \leqslant n)$ has a fixed sign (simplest assumption). The second direction gives rise to Martingales and sub/super-martingales.

The word "Martingale" has the dictionary meaning "a horse strap to keep the horse head steady". It also means a system of gambling in which the stakes are doubled or otherwise raised after each loss. In general, for us it will mean conditional zero mean process at every time point. Formally, we define the following.

Definition 12.17. *Given a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$, a sequence of r.v.s $(M_n)_{n\geqslant 0}$ is a **martingale** w.r.t. $(\mathcal{F}_n)_{n\geqslant 0}$ if 1) $M_n \in L^1(\mathcal{F}_n)$, 2) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n \ \forall\ n \geqslant 0$.*

Definition 12.18. *Given a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$, a sequence of r.v.s $(M_n)_{n\geqslant 0}$ is a **sub-martingale** w.r.t. $(\mathcal{F}_n)_{n\geqslant 0}$ if 1) $M_n \in L^1(\mathcal{F}_n)$, 2) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \geqslant M_n \ \forall\ n \geqslant 0$.*

Definition 12.19. *Given a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$, a sequence of r.v.s $(M_n)_{n\geqslant 0}$ is a **super-martingale** w.r.t. $(\mathcal{F}_n)_{n\geqslant 0}$ if 1) $M_n \in L^1(\mathcal{F}_n)$, 2) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \leqslant M_n \ \forall\ n \geqslant 0$.*

If a sequence of r.v.s is a martingale, it is both sub-martingale and super-martingale. Moreover, for a martingale $\mathbb{E}(M_n)$ is constant (increasing or decreasing for sub and super-martingales, respectively).

Example 12.20. *Given a filtration $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ and $X \in L^1(\mathcal{F})$, define $M_n = \mathbb{E}(X \mid \mathcal{F}_n)$ for $n \geqslant 0$. Then we have:*

1. *$M_n \in L^1(\mathcal{F}_n) \ \forall\ n \geqslant 0$.*
2. *$\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) = \mathbb{E}\left(\mathbb{E}(X \mid \mathcal{F}_{n+1}) \mid \mathcal{F}_n\right) = \mathbb{E}(X \mid \mathcal{F}_n) = M_n$, where the middle equality follows by Tower property.*

Example 12.21. *Suppose $X_1, X_2, \ldots$ are independent r.v.s with mean 0; and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, $M_n = X_1 + \cdots + X_n$.*

1. *$M_n \in L^1(\mathcal{F}_n)$, $\forall n \geqslant 0$.*
2. *$\mathbb{E}(M_{n+1}|\mathcal{F}_n) = \mathbb{E}(M_n + X_{n+1}|\mathcal{F}_n) = M_n + \mathbb{E}(X_{n+1}) = M_n$. Here we used that, **1)** $X \in \mathcal{G} \implies \mathbb{E}(X|\mathcal{G}) = X$; **2)** $X \perp \mathcal{G} \implies \mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.*

Example 12.22. *Given $(M_n, \mathcal{F}_n)_{n\geqslant 0}$ is a martingale, and ϕ is a convex function with $\mathbb{E}\,|\phi(M_n)| < \infty$, $\forall n$. Then $(\phi(M_n), \mathcal{F}_n)$ is a sub-martingale.*

1. $\phi(M_n) \in L^1(\mathcal{F}_n)$, $\forall n \geqslant 0$.

2. $\mathbb{E}(\phi(M_{n+1})|\mathcal{F}_n) \geqslant \phi\left(\mathbb{E}(M_{n+1}|\mathcal{F}_n)\right) = \phi(M_n)$. *(Conditional Jensen's inequality)*

Example 12.23. *Given a random vector X, uniformly distributed on a d-dimensional unit ball $B(0,1)$, and a sequence of random vectors $X_1, X_2, \dots, X_n, \dots \overset{i.i.d.}{\sim} X$, we consider the random walk*

$$S_n = \begin{cases} x_0 \in \mathbb{R}^d & \text{if } n = 0, \\ x_0 + X_1 + \cdots + X_n & \text{if } n \geqslant 1. \end{cases}$$

Let f be a super-harmonic function on $\mathbb{R}^d$, i.e., $f(x) \geqslant |B(0,1)|^{-1} \cdot \int_{y \in B(x,1)} f(y)dy$ for all x.

Then $(M_n = f(S_n))_{n \geqslant 0}$ is a super-martingale w.r.t. the filtration $\mathcal{F}_n = \sigma(X_1, \cdots, X_n), n \geqslant 0$, if $\mathbb{E}\,|M_n| < \infty$ for all $n \geqslant 0$.

1. $\phi(M_n) \in L^1(\mathcal{F}_n)$, $\forall n \geqslant 0$.

2. $\mathbb{E}(f(S_{n+1}) \mid \mathcal{F}_n) = \mathbb{E}(f(S_n + X_{n+1}) \mid \mathcal{F}_n) = \mathbb{E}(f(x + X_{n+1}))|_{x = S_n} \leqslant f(S_n) = M_n$.

Example 12.24 (Galton-Watson martingale)**.** *Let X be a non-negative integer-valued random variable with $\mathbb{E}\,X = \mu$. Let $(X_i^{(j)})$ be a sequence of iid sequence with $X_i^{(j)} \sim X$. Then set*

$$Z_0 = 1, \qquad Z_1 = X_1^{(1)}, \qquad Z_n = \sum_{i=1}^{Z_{n-1}} X_i^{(n)} \text{ for } n \geqslant 1.$$

Define

$$W_n = \frac{Z_n}{\mu^n}, \quad \mathcal{F}_n = \sigma(X_i^{(j)}, j \leqslant n) \text{ for } n \geqslant 0.$$

Notice that $W_n \in L^1(\mathcal{F}_n)$ and

$$\begin{aligned} \mathbb{E}(W_{n+1}|\mathcal{F}_n) &= \frac{1}{\mu^{n+1}}\,\mathbb{E}\left(\sum_{i=1}^{Z_n} X_i^{(n+1)} \Big| \mathcal{F}_n\right) \\ &= \frac{1}{\mu^{n+1}}\,\mathbb{E}\left(\sum_{i=1}^{\infty} X_i^{(n+1)} \mathbb{1}_{i \leqslant Z_n} \Big| \mathcal{F}_n\right) = \frac{1}{\mu^{n+1}} \sum_{i=1}^{\infty} \mathbb{1}_{i \leqslant Z_n}\,\mathbb{E}(X_i^{(n+1)}) = \frac{Z_n}{\mu^n} = W_n. \end{aligned}$$

So (W_n) is a $(\mathcal{F}_n)-$martingale.

Exercise 12.25. *Suppose there exists $\theta > 0$, such that $\mathbb{E}(\theta^X) = \theta$. Let $M_n = \theta^{Z_n}, n \geqslant 0$. Then (M_n) is a $(\mathcal{F}_n)$ martingale.*

Example 12.26 (Second moment martingale)**.** *Let (X_n) be an iid sequence of random variables with $\mathbb{E}(X_1) = 0, \mathbb{E}(X_1^2) < \infty$. Define*

$$M_n = (X_1 + ... + X_n)^2 - \sum_{i=1}^{n} \mathrm{Var}(X_i), \quad \mathcal{F}_n = \sigma(X_1, X_2, ..., X_n)$$

for $n \geqslant 1$. Then (M_n) is a $(\mathcal{F}_n)-$martingale. It is clear that $M_n \in L^1(\mathcal{F}_n)$ for any n. Notice that

$$\begin{aligned}\mathbb{E}(M_{n+1}|\mathcal{F}_n) &= \mathbb{E}((s_n + X_{n+1})^2 - \sum_{i=1}^{n+1} \mathrm{Var}(X_i) \mid \mathcal{F}_n) \\ &= s_n^2 + 2s_n\, \mathbb{E}(X_{n+1} \mid \mathcal{F}_n) + \mathbb{E}(X_{n+1}^2 \mid \mathcal{F}_n) - \sum_{i=1}^{n+1} \mathrm{Var}(X_i) \ (*) \\ &= s_n^2 - \sum_{i=1}^{n} \mathrm{Var}(X_i) = M_n\end{aligned}$$

where $s_n := \sum_{i=1}^n X_i$, and notice that at $(*)$ that following facts are applied: 1) if X is $\mathcal{G}$ measurable, then $\mathbb{E}(XY \mid \mathcal{G}) = X\, \mathbb{E}(Y \mid \mathcal{G})$, 2) if X is independent of $\mathcal{G}$, then $\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X)$.

Example 12.27 (Likelihood Ratio martingale)**.** *Let $X_n, n \geqslant 1$ be an i.i.d. sequence of random variables with $\phi(\theta) := \mathbb{E}\, e^{i\theta X_1} < \infty$. Define*

$$M_n = \phi(\theta)^{-n} \cdot e^{\theta S_n}, \quad S_n := X_1 + X_2 + \cdots + X_n, \quad \mathcal{F}_n = \sigma(X_1, X_2, ..., X_n)$$

for $n \geqslant 1$. It is clear that $M_n \in L^1(\mathcal{F}_n)$. Since $M_n \geqslant 0$ and

$$\mathbb{E}(M_{n+1} \mid \mathcal{F}_n) = \phi(\theta)^{-n-1} \cdot e^{\theta S_n}\, \mathbb{E}(e^{\theta X_{n+1}} \mid \mathcal{F}_n) = \phi(\theta)^{-n} \cdot e^{\theta S_n} = M_n,$$

Thus (M_n) is a $(\mathcal{F}_n)$-martingale.