Introduction to Martingales

1 Motivation and informal description on an exaple of Pólya's urn

In the study of random processes the following natural question arises: How does some process look at infinity? Does it converge to some fixed value? Or will it converge to a random variable maybe?

It turns out that processes known as martingales are particularly helpful in answering this. We will see this using an example of Pólya's urn.

Definition 1.1. Standard Pólya's urn is an urn that initially has 2 balls, one white and one black. Each second we pick a uniform ball from the urn and add another ball of the same colour to the urn. We denote by W_n the number of white balls in the urn when there are n balls in the urn. //

We start by quickly figuring out what the distribution of W_n is. We know that $W_2 = 1$ with probability 1. In the next step, there is a probability $1/2$ to get a white ball in which case W_3 becomes 2, otherwise it stays 1. After that, we can check that there is $1/3$ chance that $W_4 = 3$ (we had to get a white twice which has probability $1/2 \cdot 2/3 = 1/3$) and $1/3$ chance that it's 1 (we had to get two black which has the same probability as getting two white) and we are left with $1/3$ for $W_4 = 2$. We can now figure our W_5 or we can already guess that the answer is that all W_n are uniform. Therefore we get:

Proposition 1.2. *Let W_n be as in the Defintion 1.1. Then for all $n \geq 2$ we have that $W_n = \mathcal{U}[1, \ldots, n-1]$, where for any set A, $\mathcal{U}(A)$ is a uniform distribution on that set.*

Proof. Check inductively! □

Question 1.3. *For $n \geq 2$ let $X_n = \frac{W_n}{n}$ be a proportion of white balls in the Pólya's urn when there are n balls in total. Does the process X_n converge as $n \to \infty$?*

To answer the above question we first need to understand what the question is asking. The first thing we can notice is that the random variable X_n takes value in $(0, 1)$ for all n and that from Proposition 1.2 X_n is distributed uniformly on $\{\frac{1}{n}, \ldots, \frac{n-1}{n}\}$ which means that X_n could not converge to a fixed value. However, it might hold that X_n converges to some random variable. There are multiple ways in which a sequence of random variables could converge to a random variable but here we will first study the almost sure convergence.

Definition 1.4. We say that a sequence of random variables X_n on probability space $(\Omega, \sigma, \mathbb{P})$, converges almost surely (abbreviated a.s.) if

$$\mathbb{P}\Big(\lim_{n\to\infty} X_n \,\text{exists}\Big) := \mathbb{P}\Big(\Big\{\omega \in \Omega : \lim_{n\to\infty} X_n \,\text{exists}\Big\}\Big) = 1.$$

In general we say that event E holds almost surely if $\mathbb{P}(E) = 1$. Moreover, if the sequence of random variables converges almost surely then we can also define a random variable X representing the limit for which $X = \lim_{n\to\infty} X_n$ a.s. and we can, for example, define X by $X(\omega) = \lim_{n\to\infty} X_n(w)\mathbb{1}_{\{\lim_{n\to\infty} X_n(\omega)\,\text{exists}\}}$. //

Exercise 1.5. *In the setting of Defintion 1.4 check that* $\{\omega \in \Omega : \lim_{n\to\infty} X_n\,\text{exists}\}$ *is an event and that X is a well-defined random variable (i.e. check that* $\liminf$, $\limsup$ *and* $\mathbb{1}$ *are measurable functions and note that we have* $\{\omega \in \Omega : \lim_{n\to\infty} X_n\,\text{exists}\} = \{\omega \in \Omega : \limsup_{n\to\infty} X_n - \liminf_{n\to\infty} X_n = 0\}$*).*

Now that we understand some notion of convergence of random variables, we can try to see if the proportion of white balls in Pólya's urn converges almost surely. Here is a plot of outcomes of 200 experiments showing how the proportion of white balls changes for 1000 steps. In the above plot, we can see that it looks like for each experiment the

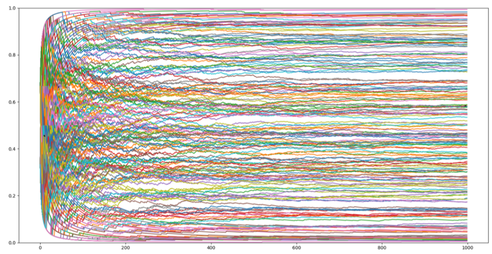

Figure 1: proportion of white balls for 200 simulations of Pólya's urn.

proportion of white balls in Pólya's urn converges but it converges to a point which seems to be equally likely anywhere in $(0, 1)$. We will now try to formally prove this convergence.

Proposition 1.6. *Let W_n be as in the Defintion 1.1, and let $X_n := \frac{W_n}{n}$ be the proportion of white balls in the Pólya's urn when there are n balls in the urn. Then X_n converges almost surely to a uniform random variable on* $(0, 1)$ *which we denote* $\mathcal{U}(0, 1)$.

We will prove the above proposition in a few separate lemmas. We start by analysing what it means for a sequence in (0,1) to converge. In particular, we can define the set $\Omega_{\text{conv}} = \{\omega \in \Omega : X_n(\omega)\,\text{converges}\}$ and we want to show that Ω_{conv} has probability 1. If $X_n(\omega)$ does not converge that could generally mean that it either goes to $\pm\infty$ (we do not have to worry about this as X_n is bounded between 0 and 1) or that it "oscillates" which we can mathematically express as

$$\exists a < b\,\text{such that}\,\forall n\,\exists\, m \geq n, M \geq n :\ X_m(\omega) < a\,\text{and}\,X_M(\omega) > b.$$

Above simply says that for some a and b with $a < b$ there are infinitely many terms in our sequence which go below a and infinitely many which are greater than b. For any fixed $a < b$ we can now define a set of $\Omega_{a,b}$ of $\omega \in \Omega$ for which $X_n(\omega)$ does not "oscillates" around a and b. Formally, $\Omega_{a,b} = \{\omega \in \Omega : \exists n\,\forall\, m \geq n\, X_m(\omega) \geq a\,\text{or}\,\forall\, m \geq n\, X_m(\omega) \leq b\}$. Now we can express Ω_{conv} in terms of $\Omega_{a,b}$ as

$$\Omega_{\text{conv}} = \bigcap_{a,b\in\mathbb{R}, a<b} \Omega_{a,b}$$

which holds as our sequence can not diverge to infinity, so it converges iff it does not oscillate around any pair or values. We would like to be able to say that if we can show that $\Omega_{a,b}$ has probability 1 for any choice of a and b that would imply that Ω has probability 1. This would be nice as $\Omega_{a,b}$ seem less complicated to control as they just tell us how the process behaves with respect to two fixed values. Unfortunately, the sequence we are taking intersection over is uncountable and we do not have in general that an uncountable intersection of probability 1 events still has probability 1. However, we might notice that this property would hold if we had countable intersections and therefore we can try to make the intersection countable. This means that we need to find a countable set of pairs (a, b) such that if $\Omega_{a,b}$ holds for all pairs in that set it also has to hold for all pairs in $\mathbb{R} \times \mathbb{R}$ with $a < b$. We can notice that for any $a < c < d < b$ we have that $\Omega_{c,d} \subset \Omega_{a,b}$, as if the sequence does not oscillate around c, d it can also not oscillate around a, b. This means that we can restrict our attention to any dense subset of $\mathbb{R}$ and as we want it to be countable a natural choice is $\mathbb{Q}$. This discussion implies the following claim.

Claim 1.7. *Let* $\Omega_{\text{conv}} = \{\omega \in \Omega : X_n(\omega) \, \text{converges}\}$ *and let the event that* X_n *does not "oscilate" around* a, b *be* $\Omega_{a,b} = \{\omega \in \Omega : \exists n \, \forall \, m \geq n \, X_m(\omega) \geq a \, \text{or} \, \forall \, m \geq n \, X_m(\omega) \leq b\}$. *Then*

$$\Omega_{\text{conv}} = \bigcap_{a,b \in \mathbb{Q}, a<b} \Omega_{a,b}$$

.

Proof. As X_n can not diverge to $\pm\infty$ the above discussion implies the claim. (If not clear write the formal proof yourself!) □

The following claim will show that it is enough to show that $\Omega_{a,b}$ has probability 1.

Claim 1.8. *For* $n \in \mathbb{N}$ *let* A_n *be events of probability* 1, *then* $\mathbb{P}(\cap_{n \in \mathbb{N}} A_n) = 1$

Proof. We want to show that intersection of sets of probability 1 still has probability 1. We can think of this as having all of our sample space except some sets of probability 0 in each A_n and then when we intersect this we want to say that we still have everything except some set of probability 0. This should hold as by union bound countable union of sets of probability 0 still has probability 0. Formally, let $A = \cap_{n \in \mathbb{N}} A_n$, using union bound we can write:

$$\mathbb{P}(A^c) = \mathbb{P}(\cup_{n \in \mathbb{N}} A_n^c) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n^c) = \sum_{n \in \mathbb{N}} (1 - \mathbb{P}(A_n)) = 0.$$

□

We can now turn to proving $\mathbb{P}(\Omega_{a,b}) = 1$. We can notice that if $\Omega_{a,b}$ holds that means that, as n increases, the number of times X_n goes below a before returning above b is finite. Therefore, to express $\Omega_{a,b}$ in terms of X_n we can use a random variable $U[a, b]$, representing the number of these "upcrossings" of $[a, b]$. In fact, if we would show that $\mathbb{E}[U[a, b]] < \infty$ this would imply $\mathbb{P}(\Omega_{a,b}) = 1$.

Definition 1.9. (Upcrossings) Let X be a random process and let $a, b \in \mathbb{R}$ with $a < b$. For a fixed $\omega \in \Omega$, by an upcrossing of $[a, b]$ by $X(\omega)$, we mean an interval of times $\{j, j+1, ..., k\}$ such that $X_j(\omega) < a$ and $X_k(\omega) > b$. We denote the number of disjoint upcrossing contained in $\{0, \ldots, n\}$ by $U_n[a, b](\omega)$ and write $U[a, b](\omega)$ for the total number of disjoint upcrossings contained in $\mathbb{Z}_+$. As $n \to \infty$, we have that $U_n[a, b]$ is increasing and converges to $U[a, b]$. //

Claim 1.10. *Let $U[a, b]$ be as in Defintion 1.9 and $\Omega_{a,b}$ as in the Claim 1.7. Then $\mathbb{E}[U[a, b]] < \infty$ implies that $\mathbb{P}(\Omega_{a,b}) = 1$.*

Proof. Suppose $\mathbb{P}\big(\Omega_{a,b}^c\big) > 0$. On the event $\Omega_{a,b}^c$ we have that the number of upcorssings is infinite as for $\omega \in \Omega_{a,b}^c$ we know that X_n "oscillates" around $[a, b]$ which means that for infinitely many time it crosses below a and then crosses above b again. Therefore, $\mathbb{E}[U[a, b]] \geq \infty \cdot \mathbb{P}\big(\Omega_{a,b}^c\big) = \infty$. □

We will now want to be able to say that if for some constant C we have that $\mathbb{E}[U_n[a, b]] < C$ then $\mathbb{E}[U[a, b]] < \infty$. This is given by the following measure theory result.

Theorem 1.11 (Monotone convergence theorem)**.** *For a sequence of non-negative measurable functions $(f_n : n \in \mathbb{N})$ such that $f_n \nearrow f$, i.e. f_n are pointwise increasing and converge to f, we have that $\mu(f) = \lim_{n\to\infty} \mu(f_n)$.*

Here you should think of $\mu(f)$ as an integral of f with respect to μ so the previous lemma tells us that for a probability measure $\mathbb{P}$ with expectation $\mathbb{E}$ for a sequence of increasing non-negative random variables U_n we have

$$\mathbb{E}\Big[\lim_{n\to\infty} U_n\Big] = \lim_{n\to\infty} \mathbb{E}[U_n] \, .$$

This gives us the following claim.

Claim 1.12. *Let $U_n[a, b]$ be as in Defintion 1.9 and $\Omega_{a,b}$ as in the Claim 1.7. Assume that there exists a constant C such that $\mathbb{E}[U_n[a, b]] \leq C$. Then we have that that $\mathbb{P}(\Omega_{a,b}) = 1$.*

Proof. The existance of constant C such that $\mathbb{E}[U_n[a, b]] \leq C$ implies, using the monotone convergence theorem, that $\mathbb{E}[U[a, b]] = \lim_{n\to\infty} \mathbb{E}[U_n[a, b]] \leq C < \infty$ (note that this limit has to exist as $\mathbb{E}[U_n[a, b]]$ is increasing (and bounded) sequence). The proof follows from Claim 1.10. □

If an event $U_n[a, b] = m$ holds that means that for $k = 1 \ldots n$, X_k went below a and after that above b and then again below a and after above b and so on for exactly m times before time n. We will say that S_1 is the first time X goes below a, T_1 the first time after S_1 that it is above b, S_2 first time after that that it is below a and so on. Formally, we have the following.

Definition 1.13. Define the stopping times S_k and T_k recursively by setting $T_0 = 0$ and for $k \geq 0$ define S_k and T_k recursively by letting

$$S_{k+1} = \inf\{t \geq T_k : X_t < a\} \quad \text{and} \quad T_{k+1} = \inf\{t \geq S_{k+1} : X_t > b\}.$$

//

Let's recall the definition of a stopping time.

Definition 1.14. For a random process X_k stopping time T is a random variable which takes values in $\mathbb{N}$ (if the process is indexed in $\mathbb{N}$) and which satisfies that $\{k < T\}$ belongs to $\sigma(X_1, \ldots X_k)$ (the smallest sigma algebra in which the random variables $X_1, \ldots, X_k$ are defined). In plain language, at the time when we sampled first k random variables we known if the time T is above or below k. //

Example 1.15. *Examples of stopping times are hitting times i.e. times when the process hits some subset of the state space, as once we sample first k values we also know for any given value if it was hit by the process by time k or not. Therefore, times in Defintion'1.13 are indeed stopping times (check this!). An example of something that is not a stopping time would be time T defined to be $\inf\{t > 0 : X_{t+5} > a\}$ for some value a, i.e. time which is defined to be 5 steps before we hit a certain region of state space. This is because when we sample $X_1, \ldots, X_k$ if T did not already happen, we still do not know that T is certainly above k, as it could happen that we do hit the wanted region at time $t + 5$. This time T would have that $T > k$ is measurable in $\sigma(X_1, \ldots, X_{k+5})$ but this can be strictly bigger sigma algebra than $\sigma(X_1, \ldots, X_k)$ (you should try to find some exact examples of the process for which the above T is not a stopping time – almost any process you pick should work).*

Lets get back to analysing $U_n[a, b]$. As $U_n[a, b] = m$ means that m upcrosssings have happened by time n, i.e. $T_m \leq n$, but $m + 1$st upcrossing has not happened, so $T_{m+1} > n$ and we have that the following events are the same:

$$\{U_n[a, b] = m\} = \{T_m \leq n < T_{m+1}\}.$$

We now want to use some property of X_n to bound $U_n[a, b]$. First, we can notice that up to time n, if event $U_n[a, b] = m$ holds X_t went m times from below a to above b which means that it travelled the distance which is at least $\geq (b - a)m$, i.e. X_t will have to travel the distance at least $(b - a)U_n[a, b]$. We now want to use the fact that we have a bound on the value of X_n and some other properties of the process itself to get an upper bound on the expectation of this distance. To do this we can notice that if $U_n[a, b] = m$ then

$$X_n = (X_n - X_{T_m}) + (X_{T_m} - X_{S_m}) + (X_{S_m} - X_{T_{m-1}}) + (X_{T_{m-1}} - X_{S_{m-1}}) + \cdots + (X_{T_1} - X_{S_1}) + X_{S_1}$$

As $X_{T_k} - X_{S_k} \geq a - b$ we have that if $U_n[a, b] = m$

$$(X_{T_m} - X_{S_m}) - \ldots - (X_{T_1} - X_{S_1}) \geq m(b - a).$$

Taking the expectation of the above seems useful as we would get an upper bound on the expectation of $U_n[a, b]$. It would be very nice if we could show something like $\mathbb{E}[X_{T_k} - X_{S_k}] \leq 0$ for most terms in the above sum, but this can not be true as we know that $X_{T_m} \geq b > a \geq X_{S_m}$. However, we have the following:

Claim 1.16. *For the process X_n from Proposition 1.6 and any stopping time T we have that $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_2] = \frac{1}{2}$, where $a \wedge b$ denotes $\min\{a, b\}$.*

Proof. We now finally need to use some special property of X_m which is that it is a sequence which satisfies that $\mathbb{E}[X_m \mid X_{m-1} = k/(m-1)] = k/(m-1)$ (as with probability $k/(m-1)$, $X_m = (k+1)/m$ and otherwise $X_m = k/m$, but $k/(m-1)\cdot(k+1)/m + (m-1-k)/(n-1)\cdot k/m = k/(m-1)$). Equivalently this means that

$$\mathbb{E}[X_m - X_{m-1} \mid X_{m-1} = k/(m-1)] = 0$$

We want to show that this implies $\mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_2]$. We can write

$$\begin{aligned} X_{T\wedge n} &= X_2 + (X_3 - X_2) + (X_4 - X_3) + \ldots + (X_{T\wedge n} - X_{T\wedge n-1}) \\ &= X_2 + \sum_{k=2}^{n-1} (X_{k+1} - X_k)\mathbb{1}_{\{k<T\}}. \end{aligned}$$

It is now enough to show that

$$\mathbb{E}\big[(X_{k+1} - X_k)\mathbb{1}_{\{k<T\}}\big] = 0$$

for any $k < n$. This holds by summing over all possible values of $X_2, \ldots, X_k$. Particularly, as T is a stopping time that means that when the values of $X_1, \ldots, X_k$ are known we also know if $\mathbb{1}_{\{k<T\}}$ is 0 or 1. If it is 0 the contribution of the term is clearly 0, but if it is 1 then we are adding a term which contains

$$\mathbb{E}[X_{k+1} - X_k \mid X_k = x_k, X_{k-1} = x_{k-1} \ldots X_3 = x_3] = \mathbb{E}[X_{k+1} - X_k \mid X_k = x_k] = 0,$$

where this holds as X_{k+1} only depends on the value of X_k and not on the past. In particular, we have that

$$\begin{aligned} &\mathbb{E}\big[(X_{k+1} - X_k)\mathbb{1}_{\{k<T\}}\big] \\ &= \sum_{\substack{x_2=\frac{1}{2} \\ x_3\in\{\frac{1}{3},\frac{2}{3}\}\ldots \\ x_k\in\{\frac{1}{k},\ldots,\frac{k-1}{k}\}}} \mathbb{E}\big[(X_{k+1} - X_k)\mathbb{1}_{\{k<T\}} \,\big|\, X_2 = x_2, \ldots X_k = x_k\big]\,\mathbb{P}(X_2 = x_2, \ldots, X_k = x_k) \\ &= \sum_{\substack{x_2=\frac{1}{2} \\ x_3\in\{\frac{1}{3},\frac{2}{3}\}\ldots \\ x_k\in\{\frac{1}{k},\ldots,\frac{k-1}{k}\}}} \mathbb{E}[(X_{k+1} - X_k) \mid X_2 = x_2, \ldots X_k = x_k]\,\mathbb{P}(X_2 = x_2, \ldots, X_k = x_k)\,\mathbb{1}_{\{k<T\}}(x_2, x_3, \ldots, x_k) \\ &= \sum_{\substack{x_2=\frac{1}{2} \\ x_3\in\{\frac{1}{3},\frac{2}{3}\}\ldots \\ x_k\in\{\frac{1}{k},\ldots,\frac{k-1}{k}\}}} \mathbb{E}[X_{k+1} - X_k \mid X_k = x_k]\,\mathbb{P}(X_2 = x_2, \ldots, X_k = x_k)\,\mathbb{1}_{\{k<T\}}(x_2, x_3, \ldots, x_k) = 0, \end{aligned}$$

where $\mathbb{1}_{\{k<T\}}(x_2, x_3, \ldots, x_k)$ is 1 if when $X_2 = x_2, \ldots, X_k = x_k$ the stopping time T does not happen by time k and it is 0 otherwise. This completes the proof. □

We can now finally prove Proposition 1.6.

Proof of Proposition 1.6. From Claim 1.12 it is enough to bound $\mathbb{E}[U_n[a,b]]$. Let S_k and T_k be defined as in Definition 1.13. If $U_n[a,b] = m$ we can write

$$X_{T_k \wedge n} - X_{S_k \wedge n} = \begin{cases} X_{T_k} - X_{S_k} \geq b - a, & \text{if } k \leq m \\ X_n - X_{S_k} \geq X_n - a, & \text{if } k = m+1 \text{ and } S_{m+1} \leq n \\ 0, & \text{otherwise.} \end{cases}$$

Summing over $k \leq n$ (as by definition of $T_k > n$ for $k > n$) we get

$$\sum_{k=1}^{n} (X_{T_k \wedge n} - X_{S_k \wedge n}) \geq (b-a)U_n[a,b] - (X_n - a)^-$$

where an expression $(x-y)^- = \max\{0, y-x\}$. From Claim 1.16 we know that $\mathbb{E}[X_{T_k \wedge n} - X_{S_k \wedge n}] = \mathbb{E}[X_2] - \mathbb{E}[X_2] = 0$ and therefore we have that

$$\mathbb{E}[U_n[a,b]] \leq \frac{\mathbb{E}[(X_n - a)^-]}{b-a}.$$

As $X_n \in (0,1)$

$$\sup_{n \in \mathbb{N}} \frac{\mathbb{E}[(X_n - a)^-]}{b-a} \leq \sup_{n \in \mathbb{N}} \frac{\mathbb{E}[|X_n| + |a|]}{b-a} \leq \frac{1+a}{b-a}$$

and this completes the proof of almost sure convergence.

We are now left with proving that the almost sure limit is indeed a uniform random variable. Let X be the almost sure limit of X_n defined by

$$X = \lim_{n \to \infty} X_n \mathbb{1}_{\{\lim_{n \to \infty} X_n \text{ exists}\}}.$$

We have that

$$\begin{aligned} \liminf_{n \to \infty} \mathbb{P}(|X_n - X| \leq \varepsilon) = \liminf_{n \to \infty} \mathbb{E}\left[\mathbb{1}_{\{|X_n - X| \leq \varepsilon\}}\right] &\geq \mathbb{E}\left[\liminf_{n \to \infty} \mathbb{1}_{\{|X_n - X| \leq \varepsilon\}}\right] \\ &\geq \mathbb{E}\left[\mathbb{1}_{\{|X_n - X| \to 0\}}\right] \\ &= \mathbb{P}\left(\lim_{n \to \infty} X_n = X\right) = 1, \end{aligned}$$

where the first inequality holds by Fatou's Lemma which we will state after this proof and the last inequality holds by the definition of X and the fact that X_n converges almost surely. This gives that $\mathbb{P}(|X_n - X| \geq \varepsilon) \to 0$ as $n \to \infty$ for any fixed ε. Therefor for any fixed $\varepsilon, \delta \in (0,1)$ there is N such that for $n \geq N$ we have that $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \delta$. So for any $c \in (0,1)$ and $n \geq N$ we have

$$\begin{aligned} \mathbb{P}(X < c) \geq \mathbb{P}(|X_n - X| \leq \varepsilon, X_n < c - \varepsilon) &= \mathbb{P}(X_n < c - \varepsilon) - \mathbb{P}(|X_n - X| \geq \varepsilon, X_n < c - \varepsilon) \\ &\geq \mathbb{P}(X_n < c - \varepsilon) - \delta \\ &\geq c - \varepsilon - \frac{1}{n} - \delta, \end{aligned}$$

where the last inequality holds from Proposition 1.2. Taking $n \to \infty$, $\delta, \varepsilon \to 0$ we get that

$$\mathbb{P}(X < c) \geq c.$$

Similarly

$$\mathbb{P}(X < c) \le \mathbb{P}(|X_n - X| \le \varepsilon, X_n < c + \varepsilon) + \mathbb{P}(|X_n - X| \ge \varepsilon) \le c + \varepsilon + \frac{1}{n} + \delta.$$

Again by taking limits

$$\mathbb{P}(X < c) \le c$$

and the proof follows. □

We now state formally Fatou's lemma we used above:

Lemma 1.17 (Fatou's Lemma). *Let $(f_n : n \in \mathbb{N})$ be a sequence of non-negative measurable functions. Then*

$$\mu(\liminf_{n\to\infty} f_n) \le \liminf_{n\to\infty} \mu(f_n).$$

The previous lemma tells us that for a probability measure $\mathbb{P}$ with expectation $\mathbb{E}$ we have that for a sequence of non-negative random variables Y_n we have

$$\mathbb{E}\Big[\liminf_{n\to\infty} Y_n\Big] \le \liminf_{n\to\infty} \mathbb{E}[Y_n]\,.$$

Exercise 1.18. *Prove Fatou's Lemma by applying the monotone convergence theorem to an appropriate function.*

We now want to investigate which properties of X_n we used to obtain that the process converges to some random variable (which is not necessarily uniform and which we do not necessarily have to be able to identify). The main ingredient of our proof was the fact that

$$\mathbb{E}[X_n \mid X_{n-1} = x_{n-1}, \dots, X_2 = x_2] = x_{n-1}$$

(in our case this was implied by only knowing what the value is when we condition on X_{n-1}, as conditional on X_{n-1} the value of X_n is independent of $X_{n-2}, \dots, X_2$, but in general this doesn't need to be true). In probability theory, there is a class of well-studied processes called martingales, and the above is, loosely speaking, the defining property of these processes.

Loosely speaking, martingales are processes which are integrable (a random variable is called integrable if $\mathbb{E}[|X|] < \infty$) and for which conditional on the past the future state of the process has expectation equal to the last known state. In the case of discrete random variables, we could think of martingales as usually being simply processes defined for times $n \in \mathbb{Z}_+(= \mathbb{N}_0)$ such that for all $n \in \mathbb{Z}_+$ we have that X_n is integrable and that for all $x_0, x_1, \dots, x_{n-1}$ we have

$$\mathbb{E}[X_n \mid X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0] = x_{n-1}.$$

This can be written in another way as

$$\mathbb{E}[X_n \mid \sigma(X_{n-1}, X_{n-2}, \dots, X_0)] = X_{n-1}$$

which should be read as whatever randomness happened until step $n-1$, the next step stays in expectation at the value of X_{n-1} (which is itself a random variable).