CHAPTER 7

INTRODUCTION TO STOCHASTIC PROCESSES

What is a stochastic process?

Definition 7.1 *Given a probability space* $(\Omega, \mathscr{F}, \mathbf{P})$*, a stochastic process is* **any** *collection of random variables defined on this probability space.*

More specifically the collection of random variables $\{X(t)\}_{t\in\mathcal{I}}$ *or alternatively written* $\{X(t) : t \in \mathcal{I}\}$*, where* $\mathcal{I}$ *is the index set. We will alternate between the notations* X_t *and* $X(t)$ *to denote the value of the stochastic process at time* t*.*

We give here the famous *R.A. Fisher* quote, who answered the same question:

> What is a stochastic process? Oh, it's just one darn thing after another.

In this chapter we start the study of stochastic processes by presenting common properties and characteristics. These properties will make the study of stochastic processes easier and they are generally desirable properties. However, it is not to be understood that all stochastic processes have these properties. In the second section of this chapter we present the easiest stochastic process we can imagine: the coin toss process (the Bernoulli process).

Probability and Stochastic Processes. By Ionuţ Florescu

7.1 GENERAL CHARACTERISTICS OF STOCHASTIC PROCESSES

7.1.1 The index set $\mathcal{I}$

The parameter that indexes the stochastic process determines the type of stochastic process we are working with.

For example if $\mathcal{I} = \{0, 1, 2 \dots\}$ (or equivalent) we obtain the so-called discrete-time stochastic processes. We will often write $\{X_n\}_{n \in \mathbb{N}}$ in this case.

If $\mathcal{I} = [0, \infty]$ we obtain the continuous-time stochastic processes. We shall write $\{X_t\}_{t>0}$ in this case.

If $\mathcal{I} = \mathbb{Z} \times \mathbb{Z}$ we may be describing a discrete random field. If $\mathcal{I} = [0, 1] \times [0, 1]$ we may be describing the structure of some random material.

These are the most common cases encountered in practice but the index set can be quite general.

7.1.2 The state space $\mathcal{S}$

This is the space where the random variables X_t which constitute our stochastic process take values. Since we are talking about random variables and random vectors, then necessarily $\mathcal{S} \subseteq \mathbb{R}$ or $\mathbb{R}^n$. Again, we have several important examples. If $\mathcal{S} \subseteq \mathbb{Z}$ we say that the process is integer valued or a process with discrete state space. If $\mathcal{S} = \mathbb{R}$ then we say that X_t is a real-valued process or a process with a continuous state space. If $\mathcal{S} = \mathbb{R}^k$ then X_t is a k-dimensional vector process.

7.1.3 Adaptiveness, filtration, standard filtration

In the special case when the index set $\mathcal{I}$ possesses a total order relationship[13] we can talk about the information contained in the process X at some moment $t \in \mathcal{I}$. To quantify this information we introduce the abstract notion of filtration.

Definition 7.2 (Filtration) *We say that a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ is a filtered probability space if and only if there exist a sequence of sigma algebras $\{\mathscr{F}_t\}_{t \in \mathcal{I}}$ included in $\mathscr{F}$ such that it is an increasing collection i.e.:*

$$\mathscr{F}_s \subseteq \mathscr{F}_t, \quad \forall s \leq t, \ s, t \in \mathcal{I}.$$

The filtration is called complete *if its first element contains all the null sets. To write this mathematically, let $\mathcal{I} = [0, \infty)$, in this case a* complete filtration *has the property that*

$$\forall N \in \mathscr{F}, \ \textit{such that } \mathbf{P}(N) = 0 \Rightarrow N \in \mathscr{F}_0.$$

From now on we assume that any filtration defined therein are complete and all filtered probability spaces are complete.

[13]i.e., for any two elements $x, y \in \mathcal{I}$, either $x \leq y$ or $y \leq x$

Definition 7.3 (Adapted stochastic process) *A stochastic process $\{X_t\}_{t\in I}$ defined on a filtered probability space $(\Omega, \mathscr{F}, \mathbf{P}, \{\mathscr{F}_t\}_{t\in\mathcal{I}})$ is called adapted if and only if X_t is $\mathscr{F}_t$-measurable for any $t \in \mathcal{I}$.*

This is an important concept. In general $\mathscr{F}_t$ quantifies the flow of information available at any moment t. By requiring that the process be adapted we insure that we can calculate probabilities related to X_t based solely on the information available at time t. Furthermore, since the filtration by definition is increasing this also says that we can calculate said probabilities at any later moment in time as well.

Remark 7.4 *On the other hand, also due to the same increasing property of filtration it may not be possible to calculate probabilities related to X_t based only on the information available in $\mathscr{F}_s$ for a moment s earlier than t ($s < t$). This is why the conditional expectation is important for stochastic processes. Recall that $\mathbf{E}[X_t|\mathscr{F}_s]$ is $\mathscr{F}_s$-measurable therefore even if we may not calculate the probabilities related to X_t we can calculate probabilities related to its best guess based on the current information $\mathbf{E}[X_t|\mathscr{F}_s]$.*

Definition 7.5 (Standard filtration) *In some cases we are only given a standard probability space (non-filtered). This usually corresponds to the case where we assume that all the information available at time t comes from the stochastic process X_t itself. No external sources of information are available. In this case we introduce the standard filtration generated by the process $\{X_t\}_{t\in\mathcal{I}}$ itself. To this end let:*

$$\mathscr{F}_t = \sigma(\{X_s : s \leq t, s \in \mathcal{I}\}),$$

where we use the notation for the sigma algebra generated by random variables given in Definition 2.13 on page 44. With this definition the collection of sigma algebras $\{\mathscr{F}_t\}_t$ is increasing and obviously the process $\{X_t\}_t$ is adapted with respect to it.

Notation. In the case when the filtration is not specified we will always construct the standard filtration and denote it with $\{\mathscr{F}_t\}_t$.

In the special case when $\mathcal{I} = \mathbb{N}$ the natural numbers, we will sometimes substitute the notation $X_1, X_2, \ldots, X_n$ instead of $\mathscr{F}_n$, as in:

$$\mathbf{E}[X_T^2|\mathscr{F}_n] = \mathbf{E}[X_T^2|X_n, \ldots, X_0]$$

7.1.4 Pathwise realizations

Suppose a stochastic process X_t is defined on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Recall that by definition for every $t \in \mathcal{I}$ fixed X_t is a random variable. On the other hand for every $\omega_0 \in \Omega$ fixed we find one realization of these variables for all t's: $X_t(\omega_0)$. Thus, for every ω_0 we find a collection of numbers representing the realization of the stochastic process - a path. Therefore, a realization of the stochastic process is a path of the Stochastic process:

$$t \mapsto X_t(\omega_0)$$

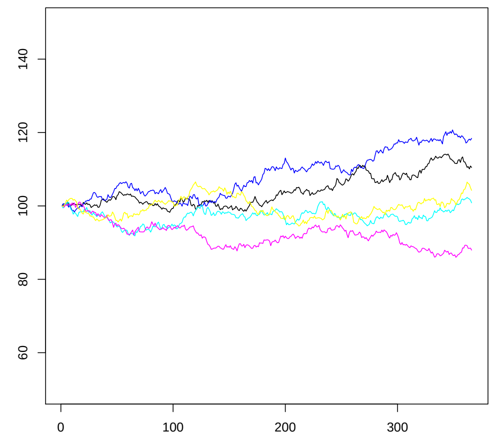

Figure 7.1 An example of 5 paths corresponding to 5 ω's for a certain stochastic process.

This means that we can identify each ω is a function from $\mathcal{I}$ into $\mathbb{R}$ and thus Ω is a subset of all the functions from $\mathcal{I}$ into $\mathbb{R}$.

In Figure 7.1 we plot five different paths each corresponding to a different realization ω. Accordingly, calculating probabilities regarding the distribution of the stochastic processes is equivalent with calculating the distribution of these paths in space. However, such a calculation is impossible when the state space is infinite or when the index set is infinite (infinite here = not countable). There is hope however.

7.1.5 The finite distribution of stochastic processes

As we have seen a stochastic process is just a collection of random variables. Thus, we have to ask: what quantities characterize a random variable? The answer is obviously its distribution. However, here we are working with a lot of variables. Depending on the number of elements in the index set $\mathcal{I}$ the stochastic process may have a finite or infinite number of components. In either case we will be concerned with the joint distribution of a finite sample taken from the process. This is due to practical consideration and the fact that in general we cannot study jointly a continuum. The processes that have a continuum structure on the set $\mathcal{I}$ serve as subject for a more advanced topic in Stochastic Differential Equations. However, even in that more advanced situation, the finite distribution of the process still constitutes a primary object of study.

Let us clarify what we mean by finite dimensional distribution. Let $\{X_t\}_{t\in\mathcal{I}}$ be a stochastic process. For any $n \geq 1$ and for any subset $\{t_1, t_2, \dots, t_n\}$ of $\mathcal{I}$ we will write $F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}$ for the joint distribution function of the variables $X_{t_1}, X_{t_2}, \dots, X_{t_n}$. The statistical properties of the process X_t are completely described by the family of distribution functions $F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}$ indexed by the n and the t_i's. This is a famous result due to Kolmogorov in the 1930's, (the exact statement is omitted – the consistency relations are very logical, you can look them up on any stochastic processes book - for example Karlin and Taylor (1975) or Øksendal (2003)).

I will restate this result again: *If we can describe these finite dimensional joint distributions we completely characterize the stochastic process.* Unfortunately, in general this is a complicated task. However, there are some properties of the stochastic processes that makes this calculation task much easier. We discuss them next.

7.1.6 Independent components

This is the most desirable property and the the most useless. Let us explain. This property implies that for any sample $\{t_1, t_2, \dots, t_n\}$ of $\mathcal{I}$ we obtain the variables $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ independent. Notice that the joint distribution $F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}$ is just the product of marginals in this case thus very easy to calculate. However, no reasonable real life processes posses this property. In effect, every new component being random implies no structure of the process so this is just a noise process. Generally speaking, in practice, this is the component in a perceived signal that one wishes to eliminate to get to the real signal process.

7.1.7 Stationary process.

A stochastic process X_t is said to be *strictly stationary* if the joint distribution functions of:

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \quad \text{and} \quad (X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$$

are the same for all $h > 0$ and any arbitrary selection $\{t_1, t_2, \dots, t_n\}$ in $\mathcal{I}$. In particular the distribution of X_t is the same for all t. Notice that this property simplifies the calculation of the joint distribution function. The condition implies that in essence the process is in equilibrium and that the particular times at which we choose to examine the process are of no relevance.

A stochastic process X_t is said to be *wide sense stationary* or *covariance stationary* if X_t has finite second moments for any t and if the covariance function $Cov(X_t, X_{t+h})$ depends only on h for all $t \in \mathcal{I}$. This is a generalization of the notion of stationarity. A strictly stationary process with finite second moments is covariance stationary. The reverse is not true, there are plenty of examples of processes which are covariance stationary but are not strictly stationary. The notion arose from real life processes that are covariance stationary but not stationary. Plus typically we can test the stationarity of covariance but it is very hard to test the strict stationarity.

Many phenomena can be described by stationary processes. Furthermore, many classes of processes which will be discussed later in this book become eventually stationary if observed for a long time.

Despite this, some of the most common processes encountered in practice – the Poisson process and the Brownian motion – are not stationary. Instead they have stationary (and independent) increments.

7.1.8 Stationary and Independent Increments

In order to talk about the increments of the process we assume that the set $\mathcal{I}$ is totally ordered.

A stochastic process X_t is said to have *independent increments* if the random variables

$$X_{t_2} - X_{t_1},\ X_{t_3} - X_{t_2},\ \ldots,\ X_{t_n} - X_{t_{n-1}}$$

are independent for any n and any choice of the sequence $\{t_1, t_2, \ldots, t_n\}$ in $\mathcal{I}$ with $t_1 < t_2 < \cdots < t_n$.

A stochastic process X_t is said to have *stationary increments* if the distribution of the random variable $X_{t+h} - X_t$ depends only on the length h of the increment and not on the time t. Notice that this is not the same as stationarity of the process itself. In fact except for the constant process there exist no process with stationary *and* independent increments which is also stationary. This is illustrated in the next Proposition.

Proposition 7.6 *If a process $\{X_t, t \in [0, \infty)\}$ has stationary and independent increments and $X_t \in L^1$, $\forall t$ then*

$$\begin{cases} \mathbf{E}[X_t] = m_0 + m_1 t \\ Var[X_t - X_0] = Var[X_1 - X_0]t, \end{cases}$$

where $m_0 = \mathbf{E}[X_0]$, and $m_1 = \mathbf{E}[X_1] - m_0$.

Proof: We will give the proof for the variance, the result for means is entirely similar (see Karlin and Taylor (1975)). Let $f(t) = Var[X_t - X_0]$. Then for any t, s we have:

$$\begin{aligned} f(t+s) &= Var[X_{t+s} - X_0] = Var[X_{t+s} - X_s + X_s - X_0] \\ &= Var[X_{t+s} - X_s] + Var[X_s - X_0] \text{ (indep. increments)} \\ &= Var[X_t - X_0] + Var[X_s - X_0] \text{ (stationary increments)} \\ &= f(t) + f(s) \end{aligned}$$

or the function f is additive (the above equation is also called Cauchy's functional equation). If we assume that the function f obeys some regularity conditions[14] then the only solution is $f(t) = f(1)t$ and the result stated in the proposition holds. ■

7.1.9 Other properties that characterize specific classes of stochastic processes

- **Point Processes.** These are special processes that count rare events. They are very useful in practice due to their frequent occurrence. For example look at

[14] these regularity conditions are either (i) f is continuous, (ii) f is monotone, (iii) f is bounded on compact intervals. In particular the third condition is satisfied by any process with finite second moments. The linearity of the function under condition (i) was first proven by Cauchy himself Cauchy (1821).

the process that gives at any time t the number of busses passing on a particular point on the 1-st street, starting from an initial time $t = 0$. This is a typical rare event ("rare" here does not refer to the frequency of the event, rather to the fact that there are gaps between event occurrence). Or look at the process that counts the number of defects in a given area of material. A particular case (and the most important) is the Poisson process which we will be studied later.

- **Markov processes.** In general terms this is a process with the property that given X_s, future values of the process (X_t with $t > s$) do not depend on any earlier X_r with $r < s$. Or, put differently the behavior of the process at any future time when its present state is known exactly is not modified by additional knowledge about its past behavior. The study of Markov processes constitutes a big part of this book. Note also that for such a process the finite distribution of the process simplifies greatly. We explain this. Using conditional distributions, for a fixed sequence of times $t_1 < t_2 < \cdots < t_n$ we can write:

$$\begin{aligned} F_{X_{t_1},X_{t_2},\dots,X_{t_n}} &= F_{X_{t_n}|X_{t_{n-1}},\dots,X_{t_1}} F_{X_{t_{n-1}}|X_{t_{n-2}},\dots,X_{t_1}} \cdots F_{X_{t_2}|X_{t_1}} F_{X_{t_1}} \\ &= F_{X_{t_n}|X_{t_{n-1}}} F_{X_{t_{n-1}}|X_{t_{n-2}}} \cdots F_{X_{t_2}|X_{t_1}} F_{X_{t_1}} \\ &= F_{X_{t_1}} \prod_{i=2}^{n} F_{X_{t_i}|X_{t_{i-1}}} \end{aligned}$$

which is a much simpler structure. In particular it means that we only need to describe one step transitions.

- **Martingales.** This is a process that has the property that the expected value of the future given the information we have today is going to be equal to the known value of the process today. These are some of the oldest processes studied in the history of probability due to their tight connection with gambling. In fact in French (the origin of the word is due to Paul Lévy) a martingale means a winning strategy (formula).

7.2 A SIMPLE PROCESS – THE BERNOULLI PROCESS

We will start the study of stochastic processes with a very simple process – tosses of a (not necessarily fair) coin. This is in fact historically the first process ever studied.

More specifically let $Y_1, Y_2, \dots$ be iid Bernoulli random variables with parameter p, i.e.,

$$Y_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

To simplify the language say a head appears when $Y_i = 1$ and a tail is obtained at the i-th toss if $Y_i = 0$. Let

$$N_k = \sum_{i=1}^{k} Y_i,$$

the number of heads up to the k-th toss, which we know is distributed as a Binomial(k, p) random variable ($N_k \sim$ Binom(k, p)).

A sample outcome may look like this:

Table 7.1 Sample Outcome

Y_i	0	0	1	0	0	1	0	0	0	0	1	1	1
N_i	0	0	1	1	1	2	2	2	2	2	3	4	5

Let S_n be the time at which n-th head (success) occurred. Mathematically:

$$S_n = \inf\{k : N_k = n\}$$

Let $X_n = S_n - S_{n-1}$ be the number of tosses to get the n-th head starting from the $(n-1)$-th head. We present a sample image below:

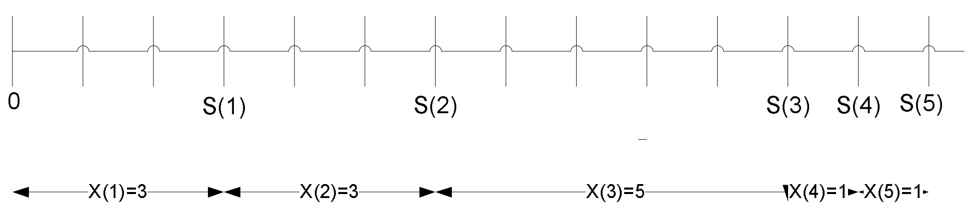

Figure 7.2 Failure and Waiting time

Proposition 7.7 *We will give some simple results about these processes.*

1) "Waiting times" $X_1, X_2 \ldots$ are iid "trials" $\sim$Geometric(p) r.v.'s.

2) The time at which the n-th head occurs is Negative Binomial, $S_n \sim$ Negative Binomial(n, p).

3) Given $N_k = n$ the distribution of $(S_1, \ldots, S_n)$ is the same as the distribution of a random sample of n numbers chosen without replacement from $\{1, 2, \ldots, k\}$.

4) Given $S_n = k$ the distribution of $(S_1, \ldots, S_{n-1})$ is the same as the distribution of a random sample of $n-1$ numbers chosen without replacement from $\{1, 2, \ldots, k-1\}$.

5) We have as sets:

$$\{S_n > k\} = \{N_k < n\}$$

6) Central Limit theorems:

$$\frac{N_k - \mathbf{E}[N_k]}{\sqrt{Var[N_k]}} = \frac{N_k - kp}{\sqrt{kp(1-p)}} \xrightarrow{D} N(0,1)$$

7)

$$\frac{S_n - \mathbf{E}[S_n]}{\sqrt{Var[S_n]}} = \frac{S_n - n/p}{\sqrt{n(1-p)/p}} \xrightarrow{D} N(0,1)$$

8) As $p \downarrow 0$

$$\frac{X_1}{\mathbf{E}[X_1]} = \frac{X_1}{1/p} \xrightarrow{D} Exponential(\lambda = 1)$$

9) As $p \downarrow 0$

$$\mathbf{P}\left\{N_{[\frac{t}{p}]} = j\right\} \longrightarrow \frac{t^j}{j!}e^{-t}$$

I will demonstrate by proving several of these properties. The rest are assigned as exercises.

For 1) and 2) there is nothing to prove since by definition X_i's are Geometric(p) random variables and S_n's are Negative Binomial. We need only to show that the X_i's are independent. See problem 7.1.

Proof (Proof for 3)): We will take $n = 4$ and $k = 100$ and prove this part only for these numbers. The general proof is identical, see problem 7.2. A typical outcome of a Bernoulli process looks like:

$$\omega : 00100101000101110000100$$

In the calculation of probability we have to have $1 \le s_1 < s_2 < s_3 < s_4 \le 100$. Using the definition of the conditional probability we can write:

$$\begin{aligned}
&\mathbf{P}(S_1 = s_1 \dots S_4 = s_4 | N_4 = 100) \\
&\quad = \frac{\mathbf{P}(S_1 = s_1 \dots S_4 = s_4 \text{ and } N_{100} = 4)}{\mathbf{P}(N_{100} = 4)} \\
&\quad = \frac{\mathbf{P}\left(\overbrace{0000\dots}^{s_1-1}1\overbrace{0000\dots}^{s_2-1}1\overbrace{0000\dots}^{s_3-1}1\overbrace{0000\dots}^{s_4-1}1\quad\overbrace{0000\dots}^{100-s_1-s_2-s_3-s_4}\right)}{\binom{100}{4}p^4(1-p)^{96}} \\
&\quad = \frac{(1-p)^{s_1-1}p(1-p)^{s_2-1}p(1-p)^{s_3-1}p(1-p)^{s_4-1}p(1-p)^{100-s_1-s_2-s_3-s_4}}{\binom{100}{4}p^4(1-p)^{96}} \\
&\quad = \frac{(1-p)^{96}p^4}{\binom{100}{4}p^4(1-p)^{96}} = \frac{1}{\binom{100}{4}}.
\end{aligned}$$

The result is significant since it means that if we only know that there have been 4 heads by the 100-th toss then any 4 tosses among these 100 are equally likely to contain the heads. ■

Proof (Proof for 8)):

$$\mathbf{P}\left(\frac{X_1}{1/p} > t\right) = \mathbf{P}\left(X_1 > \frac{t}{p}\right) = \mathbf{P}\left(X_1 > \left[\frac{t}{p}\right]\right)$$
$$= (1-p)^{\left[\frac{t}{p}\right]} = \left[(1-p)^{-\frac{1}{p}}\right]^{-p\left[\frac{t}{p}\right]} \to e^{-t},$$

since

$$\lim_{p\to 0} -p\left[\frac{t}{p}\right] = \lim_{p\to 0} -p\left(\frac{t}{p} + \left[\frac{t}{p}\right] - \frac{t}{p}\right)$$
$$= -t + \lim_{p\to 0} p \underbrace{\left(\frac{t}{p} - \left[\frac{t}{p}\right]\right)}_{\in[0,1]} = -t$$

and done since this is the tail of the exponential distribution. ■

The problems ending this chapter contain a more involved application of the Borel-Cantelli lemmas 1.21 and 1.22 to the Bernoulli process. The example is due to Dembo (2008).

Problems

7.1 Prove that the X_i's and the S_i's in Proposition 7.7 are independent.

7.2 Give a general proof of parts 3) and 4) in Proposition 7.7 for any $n, k \in \mathbb{N}$.

7.3 Show the equality of sets in part 5) of Proposition 7.7 by double inclusion.

7.4 Prove parts 6) and 7) of Proposition 7.7 by applying the Central Limit Theorem.

7.5 Prove part 9) of Proposition 7.7.

Exercises due to Amir Dembo The next problems refer to the following situation. Consider an infinite Bernoulli process with $p = 0.5$ i.e., an infinite sequence of random variables $\{Y_i, i \in \mathbb{Z}\}$ with $\mathbf{P}(Y_i = 0) = \mathbf{P}(Y_i = 1) = 0.5$ for all $i \in \mathbb{Z}$. We would like to study the length of the maximum sequence of 1's. To this end let us define some quantities.

Let

$$l_m = \max\{i \geq 1 : X_{m-i+1} = \cdots = X_m = 1\},$$

be the length of the run of 1's up to the m-th toss and including it. Obviously, l_m will be 0 if the m-th toss is a tail. We are interested in the asymptotic behavior of the longest such run from 1 to n for large n.

That is we are interested in the behavior of L_n where:

$$\begin{aligned} L_n &= \max_{m \in \{1,\dots,n\}} l_m \\ &= \max\{i \geq 1 : X_{m-i+1} = \cdots = X_m = 1, \text{ for some } m \in \{1,\dots,n\}\} \end{aligned}$$

7.6 Explain why $\mathbf{P}(l_m = i) = 2^{-(i+1)}$, for $i = 0, 1, 2, \dots$ and any m.

7.7 Apply the first Borel-Cantelli lemma 1.21 to the events $A_n = \{l_n > (1 + \varepsilon)\log_2 n\}$. Conclude that for each $\varepsilon > 0$, with probability one, $l_n \leq (1+\varepsilon)\log_2 n$ for all n large enough.
Take a countable sequence $\varepsilon_k \downarrow 0$ then conclude that:

$$\limsup_{n\to\infty} \frac{L_n}{\log_2 n} \leq 1, \quad \text{a.s.}$$

7.8 Fix $\varepsilon > 0$. Let $A_n = \{L_n < k_n\}$ for $k_n = (1-\varepsilon)\log_2 n$. Explain why

$$A_n \subseteq \bigcap_{i=1}^{m_n} B_i^c,$$

where $m_n = [n/k_n]$ (integer part) and $B_i = \{X_{(i-1)k_n+1} = \ldots = X_{ik_n} = 1$ are independent events.
Deduce that $\mathbf{P}(A_n) \leq \mathbf{P}(B_i^c)^{m_n} \leq \exp(-n^{\varepsilon}/(2\log_2 n))$, for all n large enough.

7.9 Apply the first Borel-Cantelli for the events A_n defined in problem 7.8, followed by $\varepsilon \downarrow 0$, to conclude that:

$$\liminf_{n\to\infty} \frac{L_n}{\log_2 n} \geq 1 \quad \text{a.s.}$$

7.10 Combine problems 7.7 and 7.9 together to conclude that

$$\frac{L_n}{\log_2 n} \to 1 \quad \text{a.s.}$$

Therefore the length of the maximum sequence of Heads is approximately equal to $\log_2 n$ when n, the number of tosses, is large enough.