Lecture Notes for MA 623
Stochastic Processes

Ionut Florescu

Stevens Institute of Technology
E-mail address: ifloresc@stevens.edu

2000 *Mathematics Subject Classification.* 60Gxx Stochastic Processes

Abstract.

Contents

CHAPTER 1

Probability Review.

1.1. Probability spaces. Sigma algebras.

We will use the notation from the measure theory $(\Omega, \mathscr{F}, \mathbf{P})$[1] for a probability space. Let us look to the constituent elements one at a time.

Let Ω is an abstract set. It is a set containing all the possible outcomes or results of a random experiment or phenomenon. I called it abstract because it could contain anything. For example if the experiment consists in tossing a coin once the space Ω could be represented as $\{Head, Tail\}$. However, it could just as well be represented as $\{Cap, Pajura\}$, these being the romanian equivalents of $Head$ and $Tail$. The space Ω could as well contain an infinite number of elements. For example measuring the diameter of a doughnut could result in possible numbers inside a whole range. Furthermore, measuring in inches or in centimeters would produce different albeit equivalent spaces.

We will use $\omega \in \Omega$ to denote a generic outcome or a sample point. We will use capital letters from the beginning of the alphabet A, B, C to denote events (any collection of outcomes).

We need to measure these events so we come to the next notion. The collection of events $\mathscr{F}$ represents the domain of definition for the function $\mathbf{P}$. We will need to provide internal consistencies when we define $\mathscr{F}$ to make sure that we are able to measure the information resulting from the experiment and any other event of possible interest to us. The mathematical structure for this purpose is the notion of σ-algebra (or σ-field). Before we define a σ-algebra, we will introduce a special collection of events:

$$\mathscr{P}(\Omega) = \text{The collection of all possible subsets of } \Omega = 2^{\Omega} \tag{1.1}$$

Exercise 1. Roll a die. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$. An example of a event is $A = \{$ Roll an even number$\} = \{2, 4, 6\}$. Find the cardinality (number of elements of $\mathscr{P}(\Omega)$ in this case.

[1] Sometimes (specially in statistics) the whole setup is denoted with $(S, \Sigma, \mathbf{P})$

Having defined sets we can now define operations with them: *union, intersection, complement* and slightly less important *difference and symmetric difference.*

(1.2)

$$\begin{cases} A \cup B & = \text{set of elements that are } \textbf{either} \text{ in } A \textbf{ or} \text{ in } B \\ A \cap B & = AB = \text{set of elements that are } \textbf{both} \text{ in } A \textbf{ and} \text{ in } B \\ A^c & = \bar{A} = \text{set of elements that are in } \Omega \text{ but } \textbf{not} \text{ in } A \end{cases}$$

$$\begin{cases} A \setminus B = & \text{set of elements that are in } A \text{ but } \textbf{not} \text{ in } B \\ A \triangle B = & (A \setminus B) \cup (B \setminus A) \end{cases}$$

We can of course express every operation in terms of union and intersection. There are important relations between these operations, I will stop here with the mention of the De Morgan laws:

$$\begin{cases} (A \cup B)^c & = A^c \cap B^c \\ (A \cap B)^c & = A^c \cup B^c \end{cases} \tag{1.3}$$

Definition 1.1 (Algebra on Ω)**.** A collection $\mathscr{F}$ of events in Ω is called an algebra (or field) on Ω iff:

a) $\Omega \in \mathscr{F}$
b) Closed under complementarity: If $A \subseteq \mathscr{F}$ then $A^c \subseteq \mathscr{F}$
c) Closed under finite union: If $A, B \subseteq \mathscr{F}$ then $A \cup B \subseteq \mathscr{F}$

Remark 1.2. The first two properties imply that $\varnothing \in \mathscr{F}$. The third is equivalent by de Morgan laws (1.3) with $A \cap B \subseteq \mathscr{F}$

Definition 1.3 (σ-Algebra on Ω)**.** If $\mathscr{F}$ is an algebra on Ω and in addition it is closed under countable unions then it is a σ-algebra (or σ-field) on Ω

Note: Closed under countable unions means that the property c) in Definition 1.1 is replaced with: If $n \in \mathbb{N}$ is a natural number and $A_n \subseteq \mathscr{F}$ for all n then

$$\bigcup_{n \in \mathbb{N}} A_n \subseteq \mathscr{F}.$$

From b) and c) it of course follows that the σ-algebra is also closed under countable intersection. (via De Morgan's laws)

The σ-algebra provides an appropriate domain of definition for the probability function. However, it is such an abstract thing that it will be hard to work with it. This is the reason for the next definition, it will be much easier to work with the generators of a σ-algebra. *This will be a recurring theme in probability, in order to show a property for*

a big class we show the property for a small generating set of the class and then use standard arguments to extend to the whole class.

Definition 1.4 (σ algebra generated by a class of Ω)**.** Let $\mathscr{C}$ be a collection (class) of subsets of Ω. Then $\sigma(\mathscr{C})$ is the smallest σ-algebra on Ω that contains $\mathscr{C}$. The class $\mathscr{C}$ is called the generator of the σ-algebra.

Mathematically:

(a) $\mathscr{C} \subseteq \sigma(\mathscr{C})$
(b) $\sigma(\mathscr{C})$ is a σ-field
(c) If $\mathscr{C} \subseteq \mathscr{G}$ and $\mathscr{G}$ is a σ-field then $\sigma(\mathscr{C}) \subseteq \mathscr{G}$

Remark 1.5 (Properties of σ-algebras:)**.**

- $\mathscr{P}(\Omega)$ is a σ-algebra, the largest possible σ-algebra on Ω
- If $\mathscr{F}$ is already a σ-algebra then $\sigma(\mathscr{F}) = \mathscr{F}$
- If $\mathscr{F} = \{\varnothing\}$ or $\mathscr{F} = \{\Omega\}$ then $\sigma(\mathscr{F}) = \{\varnothing, \Omega\}$, the smallest possible σ-algebra on Ω
- If $\mathscr{F} \subseteq \mathscr{F}'$ then $\sigma(\mathscr{F}) \subseteq \sigma(\mathscr{F}')$
- If $\mathscr{F} \subseteq \mathscr{F}' \subseteq \sigma(\mathscr{F})$ then $\sigma(\mathscr{F}') = \sigma(\mathscr{F})$

Remark 1.6 (Finite space Ω)**.** When the sample space is finite, we can and typically will take the sigma algebra to be $\mathscr{P}(\Omega)$. Indeed, any event of a finite space can be trivially expressed in terms of individual outcomes. In fact, if the finite space Ω contains M possible outcomes, then the number of possible events is finite and is equal with 2^M.

1.2. An Example: Borel σ-algebra.

Let Ω be a topological space (think geometry exists in this space this assures us that the open subsets exist in this space).

Definition 1.7. We define:

$$\begin{aligned} \mathscr{B}(\Omega) &= \text{The Borel } \sigma\text{-algebra} \\ &= \sigma\text{-algebra generated by the class of open subsets of } \Omega \end{aligned} \tag{1.4}$$

In the special case when $\Omega = \mathbb{R}$ we denote $\mathscr{B} = \mathscr{B}(\mathbb{R})$. $\mathscr{B}$ is *the most important* σ-algebra. The reason for that is: most experiments can be brought to equivalence with $\mathbb{R}$. Thus, if we define a probability measure on $\mathscr{B}$, we have a way to calculate probabilities for most experiments.

Most subsets of $\mathbb{R}$ are in $\mathscr{B}$. However, it is possible (though very difficult) to construct a subset of $\mathbb{R}$ explicitly which is not in $\mathscr{B}$. See [**Bil95**] page 45 for such a construction in the case $\Omega = (0, 1]$.

There is nothing special about the open sets, except for the fact that they can be defined in any topological space. In $\mathbb{R}$ we have the alternate definition which you will have to prove:

Exercise 2. Show that the following classes all generate the Borel σ-algebra, or put differently show the equality of the following collections of sets:

$$\sigma\left((a,b) : a < b \in \mathbb{R}\right) = \sigma\left([a,b] : a < b \in \mathbb{R}\right) = \sigma\left((-\infty, b) : b \in \mathbb{R}\right) \\ = \sigma\left((-\infty, b) : b \in \mathbb{Q}\right),$$

where $\mathbb{Q}$ is the set of rational numbers.

1.3. Probability Measure

We are finally in the position to define a space on which we can introduce the probability measure.

Definition 1.8 (Measurable Space.)**.** A pair $(\Omega, \mathscr{F})$, where Ω is a set and $\mathscr{F}$ is a σ-algebra on Ω is called a *Measurable Space*.

Definition 1.9 (Probability measure. Probability space)**.** Given a measurable space $(\Omega, \mathscr{F})$, a probability measure is any function $\mathcal{P} : \mathscr{F} \to [0,1]$ with the following properties:

i) $\mathcal{P}(\Omega) = 1$

ii) (countable additivity) For any sequence $\{A_n\}_{n\in\mathbb{N}}$ of disjoint events in $\mathscr{F}$ (i.e. $A_i \cap A_j = \varnothing$, for all $i \neq j$):

$$\mathcal{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

The triple $(\Omega, \mathscr{F}, \mathcal{P})$ is called a Probability Space.

The next two definitions are given for completeness, however we will use them later in this class. They are both defining more general notions than a probability measure and they will be used later in hypotheses of some theorems to show that the results apply to even more general measures than probability measures.

Definition 1.10 (Finite Measure)**.** Given a measurable space $(\Omega, \mathscr{F})$, a finite measure is a set function $\mu : \mathscr{F} \to [0,1]$ with the same countable additivity property as defined above and the measure of the space finite instead of one. More specifically the first property above is replaced with:

$$\mu(\Omega) < \infty$$

Definition 1.11 (σ-finite Measure)**.** A measure μ defined on a measurable space $(\Omega, \mathscr{F})$ is called σ-finite if it is countably additive and

there exist a partition[2] of the space Ω, $\{\Omega_i\}_{i\in I}$, and $\mu(\Omega_i) < \infty$ for all $i \in I$. Note that the index set I is allowed to be countable.

Example 1.12 (Uniform Distribution on (0,1))**.** As an example let $\Omega = (0,1)$ and $\mathscr{F} = \mathscr{B}((0,1))$. Define a probability measure U as follows: for any open interval $(a,b) \subseteq (0,1)$ let $U((a,b)) = b - a$ the length of the interval. For any other open interval O define $U(O) = U(O \cap (0,1))$.

Note that we did not specify $U(A)$ for all Borel sets A, rather only for the generators of the Borel σ-field. This illustrates the probabilistic concept presented above. In our specific situation, under very mild conditions on the generators of the σ-algebra any probability measure defined only on the generators can be uniquely extended to a probability measure on the whole σ-algebra (Carathèodory extension theorem). In particular when the generators are open sets these conditions are true and we can restrict the definition to the open sets only.

Proposition 1.13 (Elementary properties of Probability Measure)**.** *Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a Probability Space. Then:*

(1) $\forall A, B \in \mathscr{F}$ with $A \subseteq B$ then $\mathbf{P}(A) \leq \mathbf{P}(B)$
(2) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$, $\forall A, B \in \mathscr{F}$
(3) (General Inclusion-Exclusion formula, also named Poincaré formula):

$$\mathbf{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{i<j\leq n} \mathbf{P}(A_i \cap A_j) + \sum_{i<j<k\leq n} \mathbf{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^n \mathbf{P}(A_1 \cap A_2 \cdots \cap A_n) \tag{1.5}$$

Successive partial sums are alternating between over-and-under estimating.
(4) (Finite subadditivity, sometimes called Boole's inequality):

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathcal{P}(A_i), \quad \forall A_1, A_2, \ldots, A_n \in \mathscr{F}$$

1.3.1. Conditional Probability. Independence. Borel-Cantelli lemmas. Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a Probability Space. Then for $A, B \in \mathscr{F}$, with $\mathbf{P}(B) \neq 0$ we define the conditional probability of A given B as

[2] a partition of the set A is a collection of sets A_i, disjoint ($A_i \cap A_j = \emptyset$, if $i \neq j$) such that $\cup_i A_i = A$

usual by:

$$\mathbf{P}(A|\, B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

We of course also have the *chain rule formulas*:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A|\, B)\mathbf{P}(B),$$
$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A|\, B \cap C)\mathbf{P}(B|\, C)\mathbf{P}(C), \quad \text{etc.}$$

Total probability formula: Given $A_1, A_2, \ldots, A_n$ a partition of Ω (i.e. the sets A_i are disjoint and $\Omega = \bigcup_{i=1}^n A_i$), then:

$$\mathbf{P}(B) = \sum_{i=1}^{n} \mathbf{P}(B|\, A_i)\mathbf{P}(A_i), \quad \forall B \in \mathscr{F} \tag{1.6}$$

Bayes Formula: If $A_1, A_2, \ldots, A_n$ form a partition of Ω:

$$\mathbf{P}(A_j|\, B) = \frac{\mathbf{P}(B|\, A_j)\mathbf{P}(A_j)}{\sum_{i=1}^n \mathbf{P}(B|\, A_i)\mathbf{P}(A_i)}, \quad \forall B \in \mathscr{F}. \tag{1.7}$$

Definition 1.14 (Independence)**.** The events $A_1, A_2, A_3, \ldots$ are called *mutually independent* (or sometimes simply independent) if for every subset J of $\{1, 2, 3, \ldots\}$ we have:

$$\mathbf{P}\left(\bigcup_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j)$$

The events $A_1, A_2, A_3, \ldots$ are called *pairwise independent* (sometimes jointly independent) if:

$$\mathbf{P}\left(A_i \cup A_j\right) = \mathbf{P}(A_i)\mathbf{P}(A_j), \quad \forall i, j.$$

Note that jointly independent does not imply independence.

Two *sigma fields* $\mathscr{G}, \mathscr{H} \in \mathscr{F}$ *are* $\mathbf{P}$*–independent* if:

$$\mathbf{P}(G \cap H) = \mathbf{P}(G)\mathbf{P}(H), \quad \forall G \in \mathscr{G}, \forall H \in \mathscr{H}.$$

See [**Bil95**] for the definition of independence of $k \geq 2$ sigma-algebras.

1.3.2. Monotone Convergence properties of probability. Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a Probability Space.

Lemma 1.15. *The following are true:*

(i) If $A_n, A \in \mathscr{F}$ and $A_n \uparrow A$ (i.e., $A_1 \subseteq A_2 \subseteq \ldots A_n \subseteq \ldots$ and $A = \bigcup_{n \geq 1} A_n$), then: $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$ as a sequence of numbers.

(ii) If $A_n, A \in \mathscr{F}$ and $A_n \downarrow A$ (i.e., $A_1 \supseteq A_2 \supseteq \dots A_n \supseteq \dots$ and $A = \bigcap_{n\geq 1} A_n$), then: $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$ as a sequence of numbers.

(iii) (Countable subadditivity) If $A_1, A_2, \dots,$ and $\bigcup_{i=1}^{\infty} A_n \in \mathscr{F}$, with A_i's not necessarily disjoint then:

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

PROOF. (i) Let $B_1 = A_1, B_2 = A_2 \setminus A_1, \dots, B_n = A_n \setminus A_{n-1}$. Because the sequence is increasing we have that the B_i's are disjoint thus from Proposition 1.13 we obtain:

$$\mathbf{P}(A_n) = \mathbf{P}(B_1 \cup B_2 \cup \dots \cup B_n) = \sum_{i=1}^{n} \mathbf{P}(B_i).$$

Thus using countable additivity:

$$\mathbf{P}\left(\bigcup_{n\geq 1} A_n\right) = \mathbf{P}\left(\bigcup_{n\geq 1} B_n\right) = \sum_{i=1}^{\infty} \mathbf{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbf{P}(B_i) = \lim_{n\to\infty} \mathbf{P}(A_n)$$

(ii) Note that $A_n \downarrow A \Leftrightarrow {A_n}^c \uparrow A^c$ which from part (i) implies that $1 - \mathbf{P}(A_n) \uparrow 1 - \mathbf{P}(A)$.

(iii) Let $B_1 = A_1, B_2 = A_1 \cup A_2, \dots, B_n = A_1 \cup \dots \cup A_n, \dots$. From the finite subadditivity we have that $\mathbf{P}(B_n) = \mathbf{P}(A_1 \cup \dots \cup A_n) \leq \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n)$.

$\{B_n\}_{n\geq 1}$ is an increasing sequence of events, thus from (i) we get that $\mathbf{P}(\bigcup_{n=1}^{\infty} B_n) = \lim_{n\to\infty} \mathbf{P}(B_n)$. Combining the two relations above we obtain:

$$\mathbf{P}(\bigcup_{n=1}^{\infty} A_n) = \mathbf{P}(\bigcup_{n=1}^{\infty} B_n) \leq \lim_{n\to\infty} (\mathbf{P}(A_1) + \dots + \mathbf{P}(A_n)) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$$

□

Lemma 1.16. *The union of a countable number of $\mathbf{P}$-null sets is a $\mathbf{P}$-null set*

Exercise 3. Prove the above Lemma 1.16

Next we state one of the most fundamental (and useful) results in probability theory the Borel-Cantelli lemmas:

Lemma 1.17. *[The Borel-Cantelli lemmas] Let $(\Omega, \mathscr{F}, \mathcal{P})$ be a Probability Space. Let $A_1, A_2, \dots, A_n, \dots$ a sequence of events.*

First Lemma: If $\sum_{i\geq 1}\mathbf{P}(A_i)<\infty$ *then:*

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{i\geq n}A_i\right)=\mathbf{P}\left(A_i\text{'s are true infinitely often}\right)=0$$

Second Lemma: If $\sum_{i\geq 1}\mathbf{P}(A_i)=\infty$*, and in addition the events* $A_1,A_2,\ldots,A_n,\ldots$ *are independent then:*

$$\mathbf{P}\left(\bigcap_{n=1}^{\infty}\bigcup_{i\geq n}A_i\right)=\mathbf{P}\left(A_i\text{'s are true infinitely often}\right)=1$$

Let us clarify the notion of "infinitely often" and "eventually" a bit more. We say that an outcome ω happens infinitely often for the sequence $A_1,A_2,\ldots,A_n,\ldots$ if ω is in the set $\bigcap_{n=1}^{\infty}\bigcup_{i\geq n}A_i$. This means that for any n (no matter how big) there exist an $m\geq n$ and $\omega\in A_m$.

We say that an outcome ω happens eventually for the sequence $A_1,A_2,\ldots,A_n,\ldots$ if ω is in the set $\bigcup_{n=1}^{\infty}\bigcap_{i\geq n}A_i$. This means that there exist an n such that for any $m\geq n$, $\omega\in A_m$, so for an n on ω is in all such sets.

Why so complicate definitions? The basic intuition is obvious: say you roll a die infinitely many times, then it is obvious what it means for the outcome 1 to appear infinitely often. Also say the average of the rolls will eventually be arbitrarily close to 3.5. It is not so clear cut in general. The framework above provides a generalization to these notions.

Exercise 4. Show using the Cantelli lemma that when you roll a die the outcome $\{1\}$ will appear infinitely often. Also show that eventually the average of all rolls up to roll n will be within ε of 3.5 where $\varepsilon>0$ is any arbitrary real number.

1.4. Measurable Functions. Random Variables

All of these definitions with sets are consistent, however if we wish to calculate and compute numerical values related to abstract spaces we need to standardize the spaces. The first step is to give the following definitions:

Definition 1.18 (Measurable function (m.f.))**.** Let $(\Omega,\mathscr{F})$ and (S,Σ) be two measurable spaces. A function f is called measurable (function or m.f.) if and only if (notation iff) for every set $A\in\Sigma$ we have $f^{-1}(A)\in\mathscr{F}$.

1.4.1. Reduction to $\mathbb{R}$. Random variables.

Definition 1.19 (Random variable (r.v.))**.** Any measurable function with codomain $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is called a random variable.

Traditionally, the random variables are denoted with capital letters from the end of the alphabet $X, Y, Z, \ldots$ and their values are denoted with corresponding small letters $x, y, z, \ldots$.

Definition 1.20 (The distribution of a random variable)**.** Assume that on the measurable space $(\Omega, \mathscr{F})$ we define a probability measure so that it becomes a probability space $(\Omega, \mathscr{F}, \mathbf{P})$. If a random variable $X : \Omega \to \mathbb{R}$ is defined then we call its distribution, the set function μ defined on the Borel sets of $\mathbb{R}$: $\mathscr{B}(\mathbb{R})$, with values in $[0, 1]$:

$$\mu(B) = \mathbf{P}\left(\{\omega : X(\omega) \in B\}\right) = \mathbf{P}\left(X^{-1}(B)\right) = \mathbf{P} \circ X^{-1}(B)$$

Remark 1.21. First note that the measure μ is defined on sets in $\mathbb{R}$ and takes values in the interval $[0, 1]$. Therefore, the random variable X allowed us to apparently eliminate the abstract space Ω. However, this is not the case since we still have to calculate probabilities using $\mathbf{P}$ in the definition of μ above.

There is one more simplification we can make. If we use the result of the exercise 2, we see that all borel sets are generated by the same type of sets. Using the same idea as before it is enough to describe how to calculate μ for the generators. We could of course specify any type of generating sets we wish (open sets, closed sets, etc) but it turns out the simplest way is to use sets of the form $(-\infty, x]$, since we only need to specify one end of the interval (the other is always $-\infty$).

Definition 1.22. [The distribution function of a random variable] The distribution function of a random variable X is $F : \mathbb{R} \to [0, 1]$ with:

$$F(x) = \mu(-\infty, x] = \mathbf{P}\left(\{\omega : X(\omega) \in (-\infty, x]\}\right) = \mathbf{P}\left(\{\omega : X(\omega) \leq x\}\right)$$

But wait a minute, this is exactly the definition of the cumulative distribution function (cdf) you see in any lower level probability classes. It is exactly the same thing except that in an effort to dumb down (in whomever opinion it was to teach the class that way) the meaning is lost and we cannot proceed with more complicated things. From the definition above we can deduce all the elementary properties of the cdf that you have learned (right-continuity, increasing, taking values between 0 and 1). In fact let me ask you to prove this.

Exercise 5. Show that the function F in Definition 1.22 is increasing, right continuous and taking values in the interval $[0, 1]$, using proposition 1.13.

Definition 1.23 (PDF, PMF)**.** In general the distribution function F is not necessarily derivable. If it is, we call its derivative $f(x)$ the probability density function (pdf) and notice that we have in this situation:

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

Traditionally, a variable X with this property is called a continuous random variable.

Furthermore if F is piecewise constant (i.e., constant almost everywhere, or in other words there exist a countable sequence $\{a_1, a_2, \dots\}$ such that the function F is constant for every point except these a_i's) and we denote $p_i = F(a_i) - F(a_i-)$, then the collection of p_i's is the traditional probability mass function (pmf) that characterizes a discrete random variable. ($F(x-)$ is a notation for the left limit of function F at x or in a more traditional notation $\lim_{z\to x, z<x} F(z)$).

Also notice that traditional undergraduate textbooks segregate between discrete and continuous random variables. In fact there are many more and the definitions we used here cover all of them, likewise the treatment of random variables should be the same, which is now possible.

Important. So what is the point of all this? What did we just accomplish here? The answer is moving from the abstract space $(\Omega, \mathscr{F}, P)$ to something perfectly equivalent but defined on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Because of this fact we only need to define probability measures on $\mathbb{R}$ and show that things coming from our original abstract space are equivalent with these distributions on $\mathbb{R}$. We just constructed the first model for our problem.

Next we will define the simplest and one of the most important random variables.

Definition 1.24 (Indicator Function)**.** We define the indicator function of an event A as the function $\mathbf{1}_A : \Omega \to \{0, 1\}$,

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & , \quad \text{if } \omega \in A \\ 0 & , \quad \text{if } \omega \notin A \end{cases}$$

Remember this definition, it is one of the most important ones in probability. We can build on it in the following way:

Furthermore, this variable is also called the Bernoulli random variable. Notice that the variable only takes values 0 and 1 and the probability that the variable is 1 we can calculate easily using the previous definitions as being:

$$\mathbf{P} \circ \mathbf{1}_A^{-1}(\{1\}) = \mathbf{P}\{\omega : \mathbf{1}_A(\omega) = 1\} = \mathbf{P}(A).$$

Therefore the variable is distributed as a Bernoulli random variable with parameter $p = \mathbf{P}(A)$.

Definition 1.25 (Elementary (Simple) Function)**.** An elementary function g is any linear combination of the indicator functions just introduced. More specifically, there exist sets $A_1, A_2, \ldots, A_n$ all in $\mathscr{F}$ and constants $a_1, a_2, \ldots, a_n$ in $\mathbb{R}$ such that:

$$g(\omega) = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}(\omega).$$

Note that the sets A_i do not have to be disjoint but an easy exercise shows that g could be written in terms of disjoint sets.

Exercise 6. Show that any simple function g can be written as $\sum_i b_i \mathbf{1}_{B_i}$ with B_i disjoint sets (i.e. $B_i \cap B_j = \emptyset$, if $i \neq j$).

Exercise 7. Let $f : (\Omega, \mathscr{F}) \to [0, \infty]$ be a non-negative and measurable function. For all $n \geq 1$, we define:

$$g_n(\omega) := \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq f(\omega) < \}} + n\mathbf{1}_{\{f(\omega) \geq n\}} \tag{1.8}$$

(1) Show that g_n is a simple function on $(\Omega, \mathscr{F})$, for all $n \geq 1$.
(2) Show that equation (1.8) gives a partition, for all $n \geq 1$.
(3) Show that $g_n \leq g_{n+1} \leq f$, for all $n \geq 1$.
(4) Show that $g_n \uparrow f$ as $n \to \infty$[3].

1.4.2. Null element of $\mathscr{F}$. Almost sure (a.s.) statements. An event $N \in \mathscr{F}$ is called a null event if $P(N) = 0$.

Definition 1.26. A statement $\mathcal{S}$ about points $\omega \in \Omega$ is said to be true *almost surely* (a.s.), almost everywhere (a.e.) or with probability 1 (w.p.1) if the set N defined as:

$$N := \{\omega \in \Omega | \, \mathcal{S}(\omega) \text{ is true}\},$$

is in $\mathscr{F}$ and $\mathbf{P}(N) = 1$, (or N^c is a null set).

We will use the notions a.s., a.e., and w.p.1. interchangeably to denote the same thing – the definition above. For example we will say $X \geq 0$ a.s. and mean: $\mathbf{P}\{\omega | X(\omega) \geq 0\} = 1$ or equivalently $\mathbf{P}\{\omega | X(\omega) < 0\} = 0$. The notion of almost sure is a fundamental one in probability. Unlike in deterministic cases where something has

[3]This is not a.s., it is for all ω

to always be true no matter what, in probability we care about "the majority of the truth". In other words probability recognizes that some events may have extreme outcomes, but if they are extremely improbable then we do not care about them. Fundamentally, it is mathematics applied to reality.

1.4.3. Joint distribution, Random vectors. We talked about σ-algebras in the beginning and they kind of faded away after that. We will come back to them. It turns out, if there is any hope of rigorous introduction into probability and stochastic processes, they are *unavoidable*. Later, when we will talk about stochastic processes we will find out the *crucial* role they play in quantifying the information available up to a certain time. For now let us play a bit with them.

Definition 1.27 (σ-algebra generated by a random variable)**.** For a r.v. X we call the σ-algebra generated by X, denoted $\sigma(X)$ or sometime $\mathscr{F}_X$, the smallest σ-field $\mathscr{G}$ such that X is measurable on $(\Omega, \mathscr{G})$. It is the σ-algebra generated by the pre-images of Borel sets through X. Because of this we can easily show (remember that the Borel sets are generated by intervals of the type $(-\infty, \alpha]$):

$$\sigma(X) = \sigma(\{\omega | X(\omega) \leq \alpha\}, \text{ as } \alpha \text{ varies in } \mathbb{R}).$$

Similarly, given $X_1, X_2, \ldots, X_n$ random variables, we call the sigma algebra generated by them the smallest sigma algebra such that all are measurable with respect to it. It turns out we can show easily that it is the sigma algebra generated by the union of the individual sigma algebras or put more specifically $\sigma(X_i, i \leq n)$ is the smallest sigma algebra containing all $\sigma(X_i)$, for $i = 1, 2, \ldots, n$.

In the previous subsection we defined a random variable as a measurable function with codomain $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. A more specific case is when the random variable has also the domain equal to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. In this case we talk about Borel functions.

Definition 1.28 (Borel measurable function)**.** A function $g : \mathbb{R} \to \mathbb{R}$ is called Borel (measurable) function if g is a measurable function from $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ into $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Exercise 8. Show that any continuous function $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable.

Hint: Look to what happens to the preimage of sets through a continuous function.

Exercise 9. Show that any piecewise constant function is Borel measurable. (see description of piecewise constant functions in Definition 1.22

In Section 1.2 we defined Borel sigma algebras corresponding to any space Ω. We presented the special case when $\Omega = \mathbb{R}$. It really is no big deal to consider $\Omega = \mathbb{R}^n$, for some integer n, and the Borel sigma algebra generated by it. This allows us to define a random vector on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathbf{P})$ as $(X_1, X_2, \dots, X_n)$ where each X_i is a random variable. The probability $\mathbf{P}$ is defined on $\mathscr{B}(\mathbb{R}^n)$.

We can talk about its distribution (the "joint distribution" of the variables $(X_1, X_2, \dots, X_n)$) as the function:

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= \mathbf{P} \circ (X_1, X_2, \dots, X_n)^{-1} \left((-\infty, x_1] \times \cdots \times (-\infty, x_n]\right) \\ &= \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) \end{aligned}$$

We can introduce the notions of independence and joint independence using the definition in subsection 1.3.1, the probability measure $= \mathbf{P} \circ (X_1, X_2, \dots, X_n)^{-1}$ and any Borel sets. Writing more specifically it is transformed to:

Definition 1.29. The variables $(X_1, X_2, \dots, X_n, \dots)$ are independent if for every subset $J = \{j_1, j_2, \dots, j_k\}$ of $\{1, 2, 3, \dots\}$ we have:

$$\mathbf{P}\left(X_{j_1} \leq x_{j_1}, X_{j_2} \leq x_{j_2}, \dots, X_{j_k} \leq x_{j_k}\right) = \prod_{j \in J} \mathbf{P}(X_j \leq x_j)$$

1.5. Expectations of random variables.

We note that the distribution function $F(x)$ exists for any random variable. We can construct the integral with respect to F using the integration theory (details are omitted in this class) starting from indicators for which we have:

$$\mathbf{E}\left[\mathbf{1}_A\right] = \int_\Omega \mathbf{1}_A(\omega) dP(\omega) = \mathbf{P}(A)$$

In general we can construct the expectation of an integrable ($\mathbb{E}|X| < \infty$) random variable X as:

$$\mathbf{E}\left[X\right] = \int_\Omega X(\omega) dP(\omega) = \int_{-\infty}^{\infty} x dP \circ X^{-1}(x) = \int_{-\infty}^{\infty} x dF(x),$$

where we have used the transport formula (change of variable) which you can find in any graduate probability textbook. Furthermore, for any function $h : \mathbb{R} \to \mathbb{R}$ we of course can further define:

$$\mathbf{E}\left[h(X)\right] = \int_\Omega h(X(\omega)) dP(\omega) = \int_{-\infty}^{\infty} h(x) dF(x).$$

In the case when F is derivable with derivative $f(x)$ we can of course write: $dF(x) = f(x)dx$, therefore the formula reduces to the more familiar one from elementary probability classes. If F is piecewise constant then its derivative is zero a.e. and the integral reduces to a sum bringing back the formula for the expectation of a discrete random variable.

Exercise 10. Write the probability space $(\Omega, \mathscr{F}, \mathbf{P})$ for a random experiment which records the result of n independent rolls of a balanced six-sided die (including the order). Compute the expectation of the random variable $D(\omega)$ which counts the number of different sides of the die recorded during these n rolls.

The variance of a random variable X is the expectation of the function $h(x) = (x - \mu)^2$ where μ is a notation for $\mathbb{E}X$. The covariance of random variables X and Y is the expectation of the function $h(x, y) = (x - \mu_X)(y - \mu_Y)$ where again μ is a notation for the expectations of the respective random variables. The correlation is the ratio of covariance to the product of the square root of variations. More specifically:

$$\begin{aligned}
\mathbf{V}(X) &= \mathbb{E}[(x-\mu)^2] = \mathbb{E}X^2 - (\mathbb{E}X)^2 \\
\mathbf{Cov}(X,Y) &= \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y \\
\mathbf{Corr}(X,Y) &= \frac{\mathbf{Cov}(X,Y)}{\sqrt{\mathbf{V}(X)\mathbf{V}(Y)}}
\end{aligned}$$

The variable X and Y are called **uncorrelated** if the covariance (or equivalently the correlation) between them is zero. **Note** that this is not the same as the variables X and Y being independent. Independence implies that the variables are uncorrelated, however the converse is not true.

Exercise 11. Give an example of two variables X and Y which are uncorrelated but not independent.

Proposition 1.30 (Elementary properties of the expectation)**.** *The expectation has the following properties:*

(i) $\mathbf{E}[\mathbf{1}_A] = \mathbf{P}(A)$ *for any* $A \in \mathscr{F}$
(ii) If $g(\omega) = \sum_{i=1}^n a_i \mathbf{1}_{A_i}(\omega)$ *is an elementary function then* $\mathbb{E}[g] = \sum_{i=1}^n a_i \mathbf{P}(A_i)$.
(iii) If X *and* Y *are integrable r.v.'s then for any constants* α *and* β *the r.v.* $\alpha X + \beta Y$ *is integrable and* $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}X + \beta\mathbb{E}Y$.
(iv) If $X(\omega) = c$ *with probability 1 then* $\mathbb{E}X = c$.

(v) If $X \geq Y$ a.s. then $\mathbb{E}X \geq \mathbb{E}Y$. Furthermore, if $X \geq Y$ a.s. and $\mathbb{E}X = \mathbb{E}Y$ then $X = Y$ a.s.

We use the notation $L^1(\Omega)$ or sometimes just L^1 to denote the space of integrable random variables. In general:

$$L^p(\Omega) = \{X \text{ random variable s.t. } \mathbb{E}\,|X|^p < \infty\} \quad , \forall p \geq 1$$

We can make $L^p = L^p(\Omega) = L^p(\Omega, \mathscr{F}, \mathbf{P})$ a normed (metric) space by introducing the p-norm of an element (random variable) in L^p as:

$$\|X\|_p = \sqrt[p]{\mathbb{E}\left[X^p\right]}$$

1.6. Conditional Probability. Conditional Expectation.

Please read pages 5 to 9 for the definitions of conditional probability and expectation conditioned by the sigma algebra generated by a random variable.

Why do we need conditional expectation?

Conditional expectation is a fundamental concept in the theory of stochastic processes. The simple idea is the following: suppose we have no information about a certain variable then our best guess of it most of the time would be some sort of regular expectation. However, in real life it often happens that we have some partial information about the random variable (or in time we come to know more about it). Then what we should do is every time there is new information the sample space Ω or the σ-algebra $\mathscr{F}$ is changing so they need to be recalculated. That will in turn change the probability $\mathbf{P}$ which will change the expectation of the variable. The conditional expectation provides a way to recalculate the expectation of the random variable given any new "consistent" information without going through the trouble of recalculating $(\Omega, \mathscr{F}, \mathbf{P})$ every time.

It is also easy to reason that since we calculate with respect to more precise information it will be depending on this more precise information, thus it is going to be a random variable itself, "adapted" to this information.

Going back, to summarize the book notation, if X and Y are two random variables the authors define in the pages mentioned the expectation of X conditioned by the sigma-algebra generated by Y, $\sigma(Y)$ and they use the notation:

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)].$$

Note that the conditional expectation, unlike the regular expectation is a random variable measurable with respect to the sigma algebra

under which is conditioned, (in the above case with respect to $\sigma(Y)$). In general I will give you the following more general definition / theorem. We will skip the proof.

Theorem 1.31. *Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space, and let $\mathscr{K} \subseteq \mathscr{F}$ a sub-σ-algebra. Let X be a random variable on $(\Omega, \mathscr{F}, \mathbf{P})$ such that either X is positive or $X \in L^1(\Omega)$. Then there exist a random variable Y, measurable with respect to $\mathscr{K}$ with the property:*

$$\int_A Y dP = \int_A X dP \quad , \forall A \in \mathscr{K}$$

This Y is defined to be the conditional expectation of X with respect to $\mathscr{K}$ or using the notation $\mathbb{E}[X|\mathscr{K}]$.

Note that by construction Y is a $\mathscr{K}$-measurable random variable.

Proposition 1.32 (Properties of the Conditional Expectation)**.** *Let $(\Omega, \mathscr{F}, \mathbf{P})$ a probability space, and let $\mathscr{K}, \mathscr{K}_1, \mathscr{K}_2$ sub-σ-algebras. Let X and Y be random variables of the probability space. Then we have:*

(1) If $\mathscr{K} = \{\varnothing, \Omega\}$ then $\mathbb{E}[X|\mathscr{K}] = \mathbb{E}X = const.$
(2) $\mathbb{E}[\alpha X + \beta Y|\mathscr{K}] = \alpha\mathbb{E}[X|\mathscr{K}] + \beta\mathbb{E}[Y|\mathscr{K}]$ for α, β real constants.
(3) If $X \leq Y$ a.s. then $\mathbb{E}[X|\mathscr{K}] \leq \mathbb{E}[Y|\mathscr{K}]$
(4) $\mathbb{E}\left[\mathbb{E}[X|\mathscr{K}]\right] = \mathbb{E}X$
(5) If $\mathscr{K}_1 \subseteq \mathscr{K}_2$ then

$$\mathbb{E}\left[\mathbb{E}[X|\mathscr{K}_1]|\,\mathscr{K}_2\right] = \mathbb{E}\left[\mathbb{E}[X|\mathscr{K}_2]|\,\mathscr{K}_1\right] = \mathbb{E}[X|\mathscr{K}_1]$$

(6) If X is independent of $\mathscr{K}$ then

$$\mathbb{E}[X|\mathscr{K}] = \mathbb{E}[X]$$

(7) If Y is measurable with respect to $\mathscr{K}$ then

$$\mathbb{E}[XY|\mathscr{K}] = Y\mathbb{E}[X|\mathscr{K}]$$

Exercise 12. Using the Theorem-Definition 1.31 prove the seven properties of the conditional expectation in Proposition 1.32.

1.7. Generating Functions. Moment generating functions (Laplace Transform). Characteristic Function (Fourier transform)

Please read at a minimum the information in your textbook (pages 10-14) and supplement it with information from any probability textbook (including those referenced in the syllabus).

1.8. Identities. Inequalities. General Theorems

Proposition 1.33 (Jensen's Inequality)**.** *Suppose $f(\cdot)$ is a convex function, that means:*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \quad , \forall x, y \in \mathbb{R}, \forall \alpha \in [0,1].$$

Then for any integrable variable X such that $f(X) \in L^1$ we have:

$$f\left(\mathbb{E}X\right) \leq \mathbb{E}\left[f(X)\right]$$

Proof. skipped. The classical approach indicators $\rightarrow$ simple functions $\rightarrow$ positive measurable $\rightarrow$ measurable is a standard way to prove Jensen. □

Proposition 1.34 (Markov Inequality)**.** *Suppose that $g(\cdot)$ is a nondecreasing, positive measurable function. Then for any random variable X and any $\varepsilon > 0$ we have:*

$$\mathbf{P}\left(|X(\omega)| > \varepsilon\right) \leq \frac{\mathbb{E}\left[g\left(|X|\right)\right]}{g(\varepsilon)}.$$

Proof. Let $A = \{\omega : |X(\omega)| > \varepsilon\}$. We want to get to probability of A. We have using the fact that g is nonnegative:

$$\mathbb{E}\left[g\left(|X|\right)\right] = \mathbb{E}\left[g\left(|X|\right)\mathbf{1}_A\right] + \mathbb{E}\left[g\left(|X|\right)\mathbf{1}_{A^c}\right] \geq \mathbb{E}\left[g\left(|X|\right)\mathbf{1}_A\right].$$

On the set A the argument of g is greater than ε. Using this fact and that g is nondecreasing we have on A, $g\left(|X|\right) > g(\varepsilon)$. Thus we can continue:

$$\mathbb{E}\left[g\left(|X|\right)\mathbf{1}_A\right] \geq \mathbb{E}\left[g\left(\varepsilon\right)\mathbf{1}_A\right] = g(\varepsilon)\mathbf{P}(A).$$

Dividing with $g(\varepsilon)$ yields the desired result. □

Example 1.35 (Special cases of Markov Inequality)**.** These are the most common cases of the use of Markov's inequality.

(i) Take $X > 0$ a.s. and $g(x) = x$. Then we get:

$$\mathbf{P}\left(|X(\omega)| > \varepsilon\right) \leq \frac{\mathbb{E}X}{\varepsilon}$$

(ii) Take $g(x) = x^2$ and $X = Y - \mathbb{E}Y$, we then obtain:

$$\mathbf{P}\left(|Y - \mathbb{E}Y| > \varepsilon\right) \leq \frac{\mathbb{E}|Y - \mathbb{E}Y|^2}{\varepsilon^2} = \frac{\mathrm{Var}(Y)}{\varepsilon^2}.$$

A even more particular case of this is the Chebyshev's Inequality (taking $\varepsilon = k\sqrt{\mathrm{Var}(Y)} = k\sigma$).

(iii) Take $g(x) = e^{\theta x}$ for some $\theta > 0$. We get then

$$\mathbf{P}\left(X(\omega) > \varepsilon\right) \leq e^{-\theta\varepsilon}\mathbb{E}\left[e^{\theta X}\right].$$

This inequality states that the tail of the distribution decays exponentially in ε if X has finite exponential moments. With simple manipulations one can obtain Chernoff's inequality using it.

Lemma 1.36 (Cauchy-Bunyakovski-Schwarz inequality)**.** *If $X, Y \in L^2(\Omega)$, then $XY \in L^1(\Omega)$ and:*

$$|\mathbb{E}[XY]| \leq \mathbb{E}|XY| \leq \|X\|_2\|Y\|_2$$

More general we have:

Lemma 1.37 (Hölder inequality)**.** *If $1/p + 1/q = 1$, $X \in L^p(\Omega)$ and $Y \in L^q(\Omega)$ then $XY \in L^1(\Omega)$ and:*

$$\mathbb{E}|XY| \leq \|X\|_p\|Y\|_q = \left(\mathbb{E}|X|^p\right)^{\frac{1}{p}}\left(\mathbb{E}|Y|^q\right)^{\frac{1}{q}}$$

1.9. Convergence of random variables.

Asymptotic behavior is a key issue in probability theory and in the study of the stochastic processes. Why do we even need to look at the asymptotic behavior? Most of the times we cannot work with the perfect variants of the variable under study. Most of the time we will construct an approximation of the random variables (the so called model) thus it is absolutely crucial to study the conditions under which the approximation converges to the real thing. In this section we will explore the varied notions of convergence characteristic to probability theory.

1.9.1. Almost sure (a.s.) convergence. Convergence in probability. The basic notion of convergence from analysis can be translated here as a everywhere convergence. That is a sequence X_n which converges to X everywhere on the Ω or $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. For example take $X_n(\omega) = (1 - 1/n)X(\omega)$. This sequence converges to X for every omega. In general this notion is not very useful. Note that in order to have everywhere convergence we need *everywhere convergence*. It is entirely possible that the sequence X_n will converge for almost all $\omega \in \Omega$ but not for some small subset N. The point is that if this subset N has a very small probability of happening we really do not care about it. The question is how small is the probability of N and that is what differentiate the a.s. convergence from convergence in probability.

Definition 1.38 (a.s. convergence)**.** We say that X_n converges to X almost surely denoted $X_n \xrightarrow{\text{a.s.}} X$, if there exist $N \in \mathscr{F}$ with $\mathbf{P}(N) = 0$ such that $X_n(\omega) \to X(\omega)$ as $n \to \infty$ for all $\omega \in N^c$, (where N^c is a notation for the complement of the set N).

Thus here the set of omega's for which we do not have convergence have to have probability zero. Similarly with the pointwise (everywhere) convergence, the a.s. convergence is invariant with respect to continuous functionals.

Exercise 13. Show that if $f : \mathbb{R} \to \mathbb{R}$ is a continuous function and $X_n \xrightarrow{\text{a.s.}} X$, then $f(X_n) \xrightarrow{\text{a.s.}} f(X)$ as well.

A technical point here is that starting with a sequence of random variables X_n, the limiting variable may not be a random variable itself ($\mathscr{B}(\mathbb{R})$-measurable). To avoid this technical problem if one assumes that the probability space is complete (as defined next) one will always obtain random variables as the limit of random sequences (if the limit exist of course). Throughout this course we will always assume that the probability space we work with is complete.

Definition 1.39 (Complete probability space)**.** We say that the probability space $(\Omega, \mathscr{F}, \mathbf{P})$ is complete if any subset of a probability zero set in $\mathscr{F}$ is also in $\mathscr{F}$. Mathematically: if $N \in \mathscr{F}$ with $\mathbf{P}(N) = 0$, then $\forall M \subset N$ we have $M \in \mathscr{F}$.

We can easily "complete" any probability space $(\Omega, \mathscr{F}, \mathbf{P})$ by adding to its sigma-algebra all the sets of probability zero.

So that was one type of convergence (a.s.). We can make it less restrictive by looking at the measure of N and requiring that this measure instead of being zero all the time to somehow converge to zero. This is the next definition (convergence in probability).

Definition 1.40 (Convergence in probability)**.** We say that X_n converges in probability to X denoted $X_n \xrightarrow{\text{p}} X$, if the sets $N_\varepsilon(n) = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$ have the property $P(N_\varepsilon(n)) \to 0$ as $n \to \infty$, for any fixed $\varepsilon > 0$.

Theorem 1.41 (Relation between a.s. convergence and convergence in probability)**.** *We have the following relations:*

(1) If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{p} X$

(2) If $X_n \xrightarrow{p} X$ then there exist a subsequence n_k such that $X_{n_k} \xrightarrow{a.s.} X$ as $k \to \infty$

Proof. (a) Let $N^c = \{\omega : \lim |X_n(\omega) - X(\omega)| = 0\}$. We know form the definition of a.s. convergence that $P(N) = 0$.

Fix an $\varepsilon > 0$ and consider $N_\varepsilon(n) = \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}$. Let now:

$$M_k = \left(\bigcup_{n \geq k} N_\varepsilon(n) \right)^c = \bigcap_{n \geq k} N_\varepsilon(n)^c \tag{1.9}$$

- M_k's are increasing sets ($M_k = N_\varepsilon(k)^c \cap M_{k+1}$ which implies $M_k \subseteq M_{k+1}$).

- If $\omega \in M_k$ this means that for all $n \geq k$, $\omega \in N_\varepsilon(n)^c$, or $|X_n(\omega) - X(\omega)| < \varepsilon$. By definition this means that the sequence is convergent at ω, therefore $M_k \subseteq N^c$, $\forall k$, thus $\cup M_k \subseteq N^c$.

I leave it as an easy exercise to take an $\omega \in N^c$ and to show that it must exist an k_0 such that $\omega \in M_{k_0}$, therefore we will easily obtain $N^c \subseteq \cup M_k$. This will imply that $\cup M_k = N^c$ and so $\mathbf{P}(\cup M_k) = 1$, by hypothesis.

Since the sets M_k are increasing this implies that $p(M_k) \to 1$ when $k \to \infty$. Looking at the definition of M_k in (1.9) this clearly implies that

$$\mathbf{P}\left(\bigcup_{n \geq k} N_\varepsilon(n) \right) \to 0 \quad , \text{ as } k \to \infty,$$

therefore $\mathbf{P}\left(N_\varepsilon(k)\right) \to 0$, as $k \to \infty$, which is the definition of the convergence in probability.

(b) For this part we will use the Borel-Cantelli lemmas (Lemma 1.17 on page 7). We will take ε in the definition of convergence in probability of the form $\varepsilon_k > 0$ and make it to go to zero when $k \to \infty$. By the definition of convergence in probability for every such ε_k we can find an n_k, such that $\mathbf{P}\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon_k\} < 2^{-k}$, for every $n \geq n_k$. An easy process now will construct $m_k = \min(m_{k-1}, n_k)$ so that the subsequence is now increasing, while still having the above, desired property. Call:

$$N_k = \{\omega : |X_{m_k}(\omega) - X(\omega)| > \varepsilon_k\}.$$

Then from above $\mathbf{P}(N_k) < 2^{-k}$ which implies that $\sum_k \mathbf{P}(N_k) < \sum_k 2^{-k} < \infty$. Then by the first Borel-Cantelli lemma, the probability that N_k occurs infinitely often is zero. This means that with probability one N_k^c eventually. Or, the set of ω for which $\exists k_0$ and $|X_{m_k}(\omega) - X(\omega)| < \varepsilon_k\}$ for all $k \geq k_0$ has probability 1. Or the set $N := \{\omega : X_{m_k}(\omega) \to X(\omega)\}$ has probability $\mathbf{P}(N) = 1$. But this is exactly what we needed to prove. □

In general convergence in probability does not imply a.s. convergence.

Exercise 14 (Counterexample. $\xrightarrow{\text{p}}$ implies $\xrightarrow{\text{a.s.}}$)**.** You can construct your own counterexample. For instance take $\Omega = (0,1)$ with the Borel sets on it and the Lebesque measure (which is a probability measure for this Ω). Take now for every $n \in \mathbb{N}$ and $1 \leq m \leq 2^n$:

$$X_{n,m}(\omega) = \mathbf{1}_{\left[\frac{m-1}{2^n}, \frac{m}{2^n}\right]}(\omega).$$

Form a single subscript sequence by taking: $Y_1 = X_{0,1}$, $Y_2 = X_{1,1}, Y_3 = X_{1,2}$, $Y_4 = X_{2,1}, Y_5 = X_{2,2}, Y_6 = X_{2,3}, Y_7 = X_{2,4}$, etc. Draw these variables on a piece of paper for a better understanding of what is going on.

Prove that this sequence $\{Y_k\}$ has the property that $Y_k \xrightarrow{\text{p}} 0$ but $Y_k \nrightarrow Y$ a.s. In fact it does not converge for any $\omega \in \Omega$.

1.9.2. L^p convergence. Recall that we defined earlier the L^p spaces and the norm in L^p, for $p \geq 1$.

$$\|X\|_p = \sqrt[p]{\mathbb{E}\left[X^p\right]}$$

Definition 1.42. We say that the sequence X_n converges in L^p (or in the p-mean, denoted $X_n \xrightarrow{\text{L}^p(\Omega)} X$ if $X_n, X \in L^p$ and $\|X_n - X\|_p \to 0$ as $n \to \infty$ (or $\mathbb{E}(|X_n - X|^p) \to 0$ with n).

The particular case when $p = 2$ is detailed in your textbook and is called *convergence in quadratic mean*.

These L^p spaces form a complete normed vector space. This is interesting from the real analysis perspective. For our purposes the following is important:

Proposition 1.43. *Let X a random variable. Then the sequence of norms $\|X\|_p$ is non-decreasing (increasing) in p. This means that if a variable is in L^q for some q fixed then it also is in any L^r with $r \leq q$. Therefore we have (as spaces): $L^1(\Omega) \supseteq L^2(\Omega) \supseteq L^3(\Omega) \ldots$.*

PROOF. Let $p_1 > p_2$. Then the function $f(x) = |x|^{p_1/p_2}$ is convex (check this) and we can apply Jensen's inequality to the non-negative r.v. $Y = |X|^{p_2}$. The application immediately yields the desired result. □

Corollary 1.44. If $X_n \xrightarrow{\text{L}^p(\Omega)} X$ and $p \geq q$ then $X_n \xrightarrow{\text{L}^q(\Omega)} X$

PROOF. Exercise. □

Exercise 15. Show that if $X_n \xrightarrow{\text{L}^p(\Omega)} X$ then $\mathbb{E}|X_n|^p \to \mathbb{E}|X|^p$.

HINT: The $\|\cdot\|_p$ is a proper norm (recall the properties of a norm).

Next we will look into relations between the forms of convergence defined thus far.

Proposition 1.45. *If* $X_n \xrightarrow{\mathrm{L}^p(\Omega)} X$ *then* $X_n \xrightarrow{p} X$.

PROOF. This is an easy application of the Markov Inequality (Proposition 1.34). Take $g(x) = |x|^p$, and the random variable $X_n - X$. We obtain:

$$\mathbf{P}\left(|X_n - X|^p > \varepsilon\right) \leq \varepsilon^{-p}\mathbb{E}|X_n - X|^p.$$

Therefore, if $X_n \xrightarrow{\mathrm{L}^p(\Omega)} X$ then we necessarily have $X_n \xrightarrow{\mathrm{p}} X$ as well. □

Exercise 16. The converse of the previous result is not true in general. Consider the probability ensemble of Exercise 14.

Let $X_n(\omega) = n\mathbf{1}_{[0,\frac{1}{n}]}(\omega)$

Show that $X_n \xrightarrow{\mathrm{p}} X$ but $X_n \nrightarrow X$ in any L^p with $p \geq 1$.

What about convergence in L^p compared with convergence a.s.? It turns out that neither imply the other one. It is possible (easy) to come up with counterexamples for a.s. implies p-mean convergence and for p-mean convergence implies convergence a.s. However, what is true is that if both limits exist they must be the same.

Proposition 1.46. *If* $X_n \xrightarrow{\mathrm{L}^p(\Omega)} X$ *and* $X_n \xrightarrow{a.s.} Y$ *then* $X = Y$ *a.s.*

PROOF. (Sketch) We have already proven that both types of convergence imply convergence in probability. The proof then ends by showing a.s. the uniqueness of a limit in probability. □

1.9.3. Weak Convergence or Convergence in Distribution. All of the three modes of convergence discussed thus far are concerned with the case when all the variables X_n as well as their limit X are defined on the same probability space. In most applications the convergence is necessary only from the point of view of the distributions of X_n and X. I am going to stress this fact, though this is the weakest form of convergence in the sense that it is implied by all the others we are in fact discussing a totally different form of convergence.

Definition 1.47 (Convergence in Distribution – Convergence in Law – Weak-Convergence)**.** Consider a sequence of random variables X_n defined on probability spaces $(\Omega_n, \mathscr{F}_n, \mathbf{P}_n)$ (which might be all different) and a random variable X, defined on $(\Omega, \mathscr{F}, \mathbf{P})$. Let $F_n(t)$ and $F(t)$ be the corresponding distribution functions. X_n is said to converge to X in distribution (written $X_n \xrightarrow{\mathcal{D}} X$ or $F_n \Rightarrow F$) if for every point t at which F is continuous we have:

$$\lim F_n(t) = F(t).$$

Remark 1.48. There are many notations which are used interchangeably in various books, we mention $X_n \xrightarrow{\mathcal{L}} X$,$X_n \Rightarrow X$, $X_n \xrightarrow{\text{Distrib.}} X$, $X_n \xrightarrow{\text{d}} X$etc.

Remark 1.49. Why do we require t to be a continuity point of F? The simple answer is that in the discontinuity points weird things may happed even though we might have convergence everywhere else. I will give you a simple example that may illustrate this fact.

Let X_n be a $1/n$Bernoulli$(1/n)$ random variable. That is X_n takes value $1/n$ with probability $1/n$ and value 0 with probability $1 - 1/n$. Then:

$$F_n(t) = \begin{cases} 0 & \text{, if } t < \frac{1}{n} \\ 1 & \text{, if } t \geq \frac{1}{n}. \end{cases}$$

Looking at this it makes sense to say that the limit is $X = 0$ with probability 1 which has distribution function:

$$F(t) = \begin{cases} 0 & \text{, if } t < 0 \\ 1 & \text{, if } t \geq 0. \end{cases}$$

Yet, at the discontinuity point of F we have $F(0) = 1 \neq \lim F_n(0) = 0$. This is why we exclude these points from the definition.

There is one quantity where we do not care about these isolated points and that is the integral. That is why we have an alternate definition for convergence in distribution given by the next theorem. Note that it applies to random vectors X_n, X which are defined on $\mathbb{R}^d$.

Theorem 1.50. *Let X_n defined on probability spaces $(\Omega_n, \mathscr{F}_n, \mathbf{P}_n)$ and X, defined on $(\Omega, \mathscr{F}, \mathbf{P})$. Then $X_n \xrightarrow{\mathcal{D}} X$ if and only if for any bounded, continuous function on the range of X we have:*

$$\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)], \qquad \text{as } n \to \infty,$$

or equivalently:

$$\int \phi(t) dF_n(t) \to \int \phi(t) dF(t)$$

The following proposition states that (if possible to express) the convergence in probability will imply convergence in distribution. That is perhaps the reason for the name weak convergence.

Proposition 1.51. *Suppose that the sequence of random variables X_n and the random variable X are defined on the same probability space $(\Omega, \mathscr{F}, \mathbf{P})$. If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{\mathcal{D}} X$.*

Because of the fact that we are talking about apples and oranges when comparing weak convergence with anything else in general the converse of the previous theorem is false. However, there is one case where the converse is true, and that is when the limit X is a.s. a constant (notice that constants live in any probability space).

Proposition 1.52. *Let $X_n \xrightarrow{\mathcal{D}} X$ and X is a non-random constant (a.s.). Then $X_n \xrightarrow{p} X$.*

Furthermore, here is an interesting result:

Theorem 1.53 (Skorohod's representation theorem)**.** *Suppose $X_n \xrightarrow{\mathcal{D}} X$. There exists a probability space $(\Omega', \mathscr{F}', \mathbf{P}')$ and a sequence of random variables Y, Y_n on this new probability space, such that X_n has the same distribution as Y_n, X has the same distribution as Y, and $Y_n \to Y$ a.s. In other words, there is a representation of X_n and X on a single probability space, where the convergence occurs almost surely.*

Exercise 17. Write a statement explaining why the Skorohod's theorem does not contradict our earlier statement that convergence in distribution does not imply convergence a.s.

Finally, we will finish this section with the two main limit theorems from elementary probability: the law(s) of large numbers and the central limit theorem.

Theorem 1.54 (The Weak law of large numbers)**.** *Let X_n be a sequence of r.v.'s defined on probability spaces $(\Omega_n, \mathscr{F}_n, \mathbf{P}_n)$. Let us use the notations $S_n = X_1 + X_2 + \cdots + X_n$ for the sum and $\overline{X}_n = S_n/n$ for the average of the first n terms.*

Assume that X_n's are independent identically distributed (iid) with mean μ. Then $\overline{X}_n \xrightarrow{p} \mu$.

Note that the previous theorem says that this is equivalent with convergence in distribution, that is the reason for calling this result the weak law. The next result is stronger (it implies the weak law when the prob spaces are the same).

Theorem 1.55 (The Strong law of large numbers)**.** *Let X_n be a sequence of r.v.'s defined on the same probability space $(\Omega, \mathscr{F}, \mathbf{P})$. We will use the same notations from the Weak law.*

Assume that X_n's are independent identically distributed (iid) with mean μ. Then $X_n \to \mu$ a.s.

The next theorem talks about how the convergence to μ occurs.

Theorem 1.56 (The Central Limit Theorem (CLT))**.** *Let X_n be a sequence of r.v.'s defined on probability spaces $(\Omega_n, \mathscr{F}_n, \mathbf{P}_n)$. Assume as before that X_n's are iid and in addition that they have finite variance σ^2. We use the notations presented in the weak law and in addition we define the standardized variables:*

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Let Z be a $N(0,1)$ random variable. Then we have:

$$Z_n \xrightarrow{\mathcal{D}} Z.$$

1.10. Uniform Integrability[4]

We have seen that convergence a.s. and convergence in L^p are generally not compatible. However, we will give next an integrability condition that together with convergence in probability will imply convergence in p-mean.

Definition 1.57 (Uniform Integrability criterion)**.** A collection of random variables $\{X_\alpha\}_{\alpha\in\mathcal{I}}$ is called uniform integrable (U.I.) if:

$$\lim_{M\to\infty} \sup_{\alpha} \mathbb{E}\left[|X_\alpha| \mathbf{1}_{\{|X_\alpha|>M\}}\right] = 0.$$

In other words the tails of the expectation converge to 0 uniformly for all the family.

Theorem 1.58. *If $X_n \xrightarrow{p} X$ and for a fixed $p \geq 1$ the family $\{|X_n|^p\}_{n\in\mathbb{N}}$ in U.I. then $X_n \xrightarrow{p} X$*

For the proof see [GS01, Theorem 7.10.3]
We will give a few more details about U.I.

Example 1.59. Examples of U.I. families:

- Any r.v. $X \in L^1$ is U.I.
  ($\mathbb{E}|X| < \infty$ implies immediately $\mathbb{E}\left[|X|\mathbf{1}_{\{|X|>M\}}\right] \xrightarrow[M\to\infty]{} 0$)
- Let the family X_α bounded by an integrable random variable i.e., $|X_\alpha| \leq Y$ and $Y \in L^1$ then X_α is U.I.
  Indeed, we have $\mathbb{E}\left[|X_\alpha|\mathbf{1}_{\{|X_\alpha|>M\}}\right] \leq \mathbb{E}\left[Y\mathbf{1}_{\{|Y|>M\}}\right]$, which does not depend on α and converges to 0 with M as in the previous example.

[4]Not normally taught in Ma611

- Any finite collection of r.v.'s in L^1 is U.I.
  This is just an application of the previous point. If $\{X_1, X_2, \ldots, X_n\}$ is the collection of integrable r.v.'s take for example $Y = |X_1| + |X_2| + \cdots + |X_n|$.
- The family $\{a_\alpha Y\}$ with $Y \in L^1$ and $a_\alpha \in [-1, 1]$, non-random constants is U.I.
- Any bounded collection of integrable r.v.'s is U.I.

Next we give a very useful criterion for U.I.

Proposition 1.60. *A family of r.v.'s $\{X_\alpha\}_{\alpha\in\mathcal{I}}$ is uniform integrable if $\mathbb{E}f(|X_\alpha|) \le C$ for some finite C and all α, where $f \ge 0$ is any function such that $f(x)/x \to \infty$ as $x \to \infty$.*

Here is an example of a family which is not U.I.

Example 1.61. Let us consider the probability space of all infinite sequences of coin tosses (we will see this space later on in reference to Bernoulli process). Assume that the coin is fair.

Let $X_n = \inf_{i>n}\{\text{toss } i \text{ is a } H\}$, the first toss after n where we obtain a head. Then for any M we can find $n \ge M$ therefore $X_n > n \ge M$ thus $\mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n|>M\}}\right] = \mathbb{E}[X_n] > n$, implying that X_n is not U.I.

1.11. Exchanging the order of limits and expectations

This is an important question. In many cases we need to put the limit under the integral sign, but are we doing it correctly?

There are 4 results that can help you with this question.

The first two results results basically require the sequence and the limit to be integrable.

Theorem 1.62 (Dominated Convergence)**.** *If there exists a random variable Y such that $\mathbb{E}Y < \infty$, $X_n \le Y$ for all n and if we have $X_n \xrightarrow{p} X$, then $\mathbb{E}X_n \to \mathbb{E}X$ as well.*

In the particular case when Y is non-random we obtain:

Corollary 1.63 (Bounded Convergence)**.** Suppose that $X_n \le C$, $\forall n$ for some finite constant C. If $X_n \xrightarrow{\text{p}} X$, then $\mathbb{E}X_n \to \mathbb{E}X$ as well.

In the case of monotone (increasing) convergence of non-negative r.v.'s we can exchange the limit and the expectation even if X is *non-integrable*.

Theorem 1.64 (Monotone Convergence)**.** *If $X_n \ge 0$ and $X_n(\omega) \uparrow X(\omega)$ a.s. then $\mathbb{E}X_n \uparrow \mathbb{E}X$. This is true even if $X(\omega) = \infty$ for some $\omega \in \Omega$*

Remark 1.65. You may think that as we have increasing convergence we must also have decreasing convergence. We indeed have but the result is not that useful. It requires the extra assumption $\mathbb{E}[X_1] < \infty$. But, if we make this assumption the exchange of limit and integral is true already from the dominated convergence theorem. If we wish to drop the extra assumption the result is no longer true as the next example demonstrates.

Example 1.66. Let Z be a random variable such that $\mathbb{E}Z = \infty$. Take $X_1 = Z$, and in general $X_n(\omega) = n^{-1}Z(\omega)$. Then we have that $\mathbb{E}X_n = \infty$, for any n but $X_n \downarrow 0$ wherever Z is finite.

Practice your understanding solving the following exercise:

Exercise 18. Let Y_n a sequence of non-negative random variables. Use the Monotone Convergence Theorem to show that:

$$\mathbb{E}\left[\sum_{n=1}^{\infty} Y_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[Y_n].$$

Continue by showing that if $X \geq 0$ a.s. and A_n are disjoint sets with $\mathbf{P}(\cup_n A_n) = 1$ (partition of Ω), then:

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{E}(X\mathbf{1}_{A_n}).$$

Furthermore, show that the result applies also when $X \in L^1$.

The last result presented bellow is the most useful in practice; we do not require the sequence or the limit to be integrable nor do we require a special (monotone) form of convergence. We only require the existence of a lower bound. However, the result is restrictive, it only allows exchange of the lim inf with the expectation.

Lemma 1.67 (Fatou's Lemma)**.** *Suppose that X_n is a sequence of random variables such that there exist a $Y \in L^1$ with $X_n > Y$ for all n. Then we have:*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]$$

Here:

$$\liminf_{n\to\infty} X_n = \lim_{n\to\infty} \left\{ \inf_{k\geq n} X_k \right\}.$$

CHAPTER 2

Introduction to Stochastic Processes

What is a stochastic process?

Definition 2.1. 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 alternately use X_t to denote $X(t)$.

We will give here the famous *R.A. Fisher* quotation:

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

We will next describe some characteristics important for all stochastic processes.

2.1. General characteristics of Stochastic processes

2.1.1. The index parameter $\mathcal{I}$. The parameter that indexes the stochastic process determines the type of stochastic process we are working with.

For example if $\mathcal{I} = \{0, 1, 2 \ldots\}$ (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.

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.

2.1.2. The state space $\mathcal{S}$. This is the space where the random variables X_t which constitute our stochastic process take values. Again we have several important examples. If $\mathcal{S} \subseteq \mathbb{Z}$ we say that the process is integer valued or a discrete state process. If $\mathcal{S} = \mathbb{R}$ then we say that X_t is a real-valued process. $\mathcal{S} = \mathbb{R}^k$ then X_t is a k-vector process.

2.1.3. The finite distribution of stochastic processes. As we have seen a stochastic process is just a collection of random variables. Thus we have to look into what quantities characterizes a random variable. That is obviously its distribution. However, here we are working with a lot of them. Depending on the index set $\mathcal{I}$ the stochastic process may be finite or infinite. In either case we will be primarily 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 $\mathcal{I}$ serves as subject for a later course in Stochastic Differential equations. However, even in that case the finite distribution of the process serves as a primary object of study.

More specifically, let $\{X_t\}_{t\in\mathcal{I}}$ be a stochastic process. For any subset $\{t_1, t_2, \ldots, t_n\}$ of $\mathcal{I}$ we will write $F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}$ for the joint distribution function of the variables $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$. The statistical properties of the process X_t are completely described by the family of distribution functions $F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}$ 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 page 9 of [**KT75**]).

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

2.1.4. Independent components of the process. This is the most desirable property and the the most useless. Let us explain. This property implies that for any sample $\{t_1, t_2, \ldots, t_n\}$ of $\mathcal{I}$ we get $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ independent. Notice that the joint distribution $F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}$ 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 one wishes to eliminate to get to the real signal process.

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

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

are the same for all $h > 0$ and any arbitrary selection $\{t_1, t_2, \ldots, t_n\}$ 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. There are 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.

Many phenomena can be described by stationary processes. We will discuss them later in this course. However, 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.

2.1.6. Stationary and Independent Increments. 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 choice of the sequence $\{t_1, t_2, \ldots, t_n\}$ in $\mathcal{I}$ with $t_1 < t_2 < \cdots < t_n$. Notice that we are talking about an order on the set $\mathcal{I}$ so an ordering relation must be defined prior to talking about increments.

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.

Proposition 2.2. *If a process $\{X_t, t \in [0, \infty)\}$ has stationary independent increments and $X_t \in L^1$, $\forall t$ then*

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

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

Proof. We will indicate the proof only for the variances, the result for means you can read in [**KT75**, page 28]. Also note the error in the

statement in the book. Let $f(t) = Var[X_t - X_0]$. Then for any t, s we have:

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

Using the result in the textbook [**KT75**, page 28] we obtain the solution $f(t) = f(1)t$ and the result stated in the proposition. □

2.1.7. Other properties that characterize specific classes of stochastic processes.

- **Markov processes.** In general terms this is a process with the property that given X_s, the 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 concerning its past behavior. The study of Markov processes constitutes a big part of this class. Note also that for such a process the finite distribution of the process simplifies greatly.
- **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. We will study Martingales later in this class.
- **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 the process that gives at any time t the number of busses passing on Washington street and 6^{th} 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 study in this class.

2.2. A Simple process – The Bernoulli process

We will start by studying a very simple process – tosses of a (not necessarily fair) coin. More specifically let $Y_1, Y_2, \ldots$ be iid Bernoulli

random variables with parameter p, i.e.,

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

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 you know is distributed as a Binomial(k, p) random variable. We will use the notation $N_k \sim$ Binomial(k, p) from now on to denote distribution of random variables.

A sample outcome may look like this:

TABLE 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:

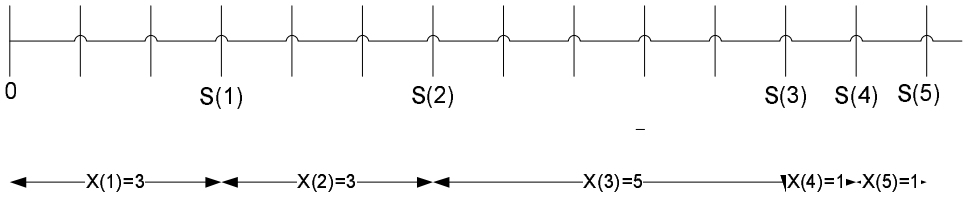

FIGURE 1. Failure and Waiting time

Proposition 2.3. *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:

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

7)

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

8) As $p \downarrow 0$

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

9) As $p \downarrow 0$

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

Exercise 19. Prove the previous properties. To make it a bit easier in the parts 3 and 4, take $n = 4$ and $k = 100$ (The general proof is identical).

I will give you some hints. For 1) there is nothing to prove for X_i's are Geometric(p) random variables. You need only to show that they are independent. The solution for 3 I have already written so there it is.

Proof for 3). A typical outcome of a Bernoulli process looks like:

$$\omega : 00100101000101110000100$$

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

write:

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

□

Proof for 8).

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

since

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

□

We will finish this chapter with a more involved application of the Borel-Cantelli lemma 1.17 to the Bernoulli process.

Exercise 20 (Due to Amir Dembo)**.** 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 $l_m = \max\{i \geq 1 : X_{m-i+1} = \dots = X_m = 1\}$, 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 the behavior of L_n where:

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

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

(b) Applying the first Borel-Cantelli lemma 1.17 for $A_n = \{l_n > (1+\varepsilon)\log_2 n\}$ show that for each$\varepsilon > 0$, with probability one, $l_n \leq (1+\varepsilon)\log_2 n$ for all n large enough. Considering a countable sequence $\varepsilon_k \downarrow 0$ conclude that:

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

(c) 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.

(d) Applying the first Borel-Cantelli for the events A_n of part (c), followed by $\varepsilon \downarrow 0$, conclude that:

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

(e) Putting (b) and (d) together we 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.

CHAPTER 3

The Poisson process

I believe the treatment of the Poisson process is absolutely essential in modern stochastic processes treatment due to the vast array of applications of this process. We will start with basic definitions first.

3.1. Definitions.

Definition 3.1 (Counting Process)**.** N_t is a counting process if and only if

(1) $N_t \in \{0, 1, 2 \ldots\}, \forall t$
(2) N_t is non decreasing as a function of t

Here N_t is non-decreasing means that all the sample path $N_t(w)$ are non-decreasing as a function of t for every $w \in \Omega$, w fixed.

Definition 3.2 (Poisson Process)**.** $N(t)$ is a Poisson(λ) process if it is a counting process and in addition

(1) $N(0) = 0$.
(2) $N(t)$ has stationary independent increments.
(3) $P(N(h) = 1) = \lambda h + o(h)$.
(4) $P(N(h) \geq 2) = o(h)$.

Facts:

1: $f \sim o(g)$ if and only if $\lim_{x \to 0} \frac{f(x)}{g(x)} = 0$

2: $f \sim O(g)$ is and only if there exist c_1, c_2 constants, such that $c_1 \leq \frac{f(x)}{g(x)} \leq c_2$, $\forall x$ in a neighborhood of 0.

Theorem 3.3. *If $N(t)$ is a Poisson(λ) process then*[1]

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Proof. A standard proof (presented in the notes handed in class) derive and solves the Kolmogorov's forward differential equations of the Poisson(λ) process, a discrete state space Markov Chain. This method

[1]Note that $P(N(t) = n) = P(N(s + t) - N(s))$ by the stationarity of the increments.

will be seen later and it is worth your time to read and understand that proof as well.

Here we will approach the proof a bit different. The idea is to approximate a Poisson(λ) process with a Bernoulli process and then pass to the limit.

Fix $t > 0$, cut $[0, t]$ into 2^k equally spaced intervals.

Let $\widetilde{N}_k$ = the number of these 2^k intervals with at least one event in them. Note that we have the condition

$$\widetilde{N}_k \leq N_t \quad \begin{cases} = & \text{only when each interval contain at most 1 event} \\ \leq & \text{always} \end{cases}$$

In addition, let

$$E_k = \{N_t > \widetilde{N}_k\} = \bigcup_{i=0}^{2^{k-1}} \left\{ \underbrace{N\left(\frac{i+1}{2^k}t\right) - N\left(\frac{i}{2^k}t\right) \geq 2}_{\text{at least 1 interval with 2 events}} \right\}$$

Take probability on both sides

$$\begin{aligned} P(E_k) &\leq \sum_{i=0}^{2^{k-1}} P\left(N\left(\frac{i+1}{2^k}t\right) - N\left(\frac{i}{2^k}t\right) \geq 2\right) \\ &= \sum_{i=0}^{2^{k-1}} P\left(N\left(\frac{1}{2^k}t\right) \geq 2\right) \qquad \text{(by stationarity)} \\ &= 2^k o\left(\frac{t}{2^k}\right) = \frac{o(t/2^k)}{t/2^k}\, t \xrightarrow{k\to\infty} 0, \end{aligned}$$

for every t fixed. So $\mathbf{P}(E_k) \xrightarrow{k\to\infty} 0$

Now $\widetilde{N}_k \sim$ Binomial$(2^k, \lambda\frac{t}{2^k} + 2o(\frac{t}{2^k}))$ Note that $\lambda\frac{t}{2^k} + 2o(\frac{t}{2^k})$ is the probability that at least one event occurs in an interval, i.e., $p = \mathbf{P}(N(\frac{t}{2^k}) = 1) + \mathbf{P}(N(\frac{t}{2^k}) \geq 2)$

Exercise 21. If $W_k \sim$ Binomial(k, p_k) and $kp_k \to \lambda$ when $k \to \infty$, then

$$W_k \xrightarrow{\mathcal{D}} \text{Poisson}(\lambda)$$

,i.e.

$$P(W_k = n) \to \frac{\lambda^n}{n!} e^{-\lambda}$$

In our case

$$2^k\left(\lambda\frac{t}{2^k} + 2o(\frac{t}{2^k})\right) = \lambda t + 2\frac{o(t/2^k)}{t/2^k} t \xrightarrow{k\to\infty} \lambda t$$

Therefore the Exercise 21 implies

$$\widetilde{N}_k \xrightarrow{\mathcal{D}} \text{Poisson}(\lambda t) \tag{3.1}$$

OR

$$P(\widetilde{N}_k = n) \to \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Furthermore, $N(t) = \widetilde{N}_k + \underbrace{N(t) - \tilde{N}_k}$ and we know $P(\widetilde{N}_k \neq N(t)) = P(E_k) \to 0$ as $k \to \infty$. Therefore, we must have:

$$\widetilde{N}_k \xrightarrow{\mathcal{D}} N(t) \tag{3.2}$$

From (3.1) and (3.2) the limits of $\widetilde{N}_k$ must be the same thing. Done. □

3.2. Inter-arrival and waiting time

Let:

$$\begin{aligned} X_1 &= \text{time of the first event} \\ X_2 &= \text{time between 1st and 2nd event} \\ &\vdots \\ X_n &= \text{time between (n-1)-th and n-th event} \end{aligned}$$

Let $S_i =$ time of the i-th event and notice that:

$$\begin{aligned} S_i &= \sum_{j=1}^{i} X_j \\ S_n &= \inf\{t : N(t) \geq n\} = \inf\{t : N(t) = n\} \end{aligned}$$

Proposition 3.4. *$X_1, X_2 \ldots$ are iid random variable, exponentially distributed with mean $\frac{1}{\lambda}$.*

We will not prove this proposition instead we will prove the following claim:

Claim–Evidence: The distribution of S_n is Gamma(n, λ) or the p.d.f. of S_n is given by:

$$f_{S_n}(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!} \qquad t \geq 0$$

We note that the exponential distribution is a special case of Gamma distribution. In fact, Exponential$(\lambda) =$ Gamma$(1, \lambda)$. This is a useful fact to know since the Gamma distribution has some nice property, one of them being that if the two variables added are independent then:

$$\text{Gamma}(\alpha_1, \beta) + \text{Gamma}(\alpha_2, \beta) \overset{\mathcal{D}}{=} \text{Gamma}(\alpha_1 + \alpha_2, \beta)$$

For this reason if the Proposition 3.4 is true then we must have the distribution of the arrival times S_n as:

$$\begin{aligned} S_1 &= X_1 \sim \text{Gamma}(1, \lambda) \\ S_2 &= X_1 + X_2 \sim \text{Gamma}(2, \lambda) \\ &\vdots \\ S_n &= X_1 + \ldots + X_n \sim \text{Gamma}(n, \lambda), \end{aligned}$$

Therefore, proving the claim adds evidence in favor of the Proposition 3.4. In fact we will prove the Proposition using the claim.

PROOF OF THE CLAIM-EVIDENCE. We know that: $\{S_n \leq t\} = \{N(t) \geq n\}$ (convince yourself of the truth of this affirmation). Thus the c.d.f

$$F_{S_n}(t) = \mathbf{P}\{S_n \leq t\} = \mathbf{P}\{N(t) \geq n\} = \sum_{j=n}^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t}$$

Take the derivative with respect to t: $\frac{\partial}{\partial t}$

$$\begin{aligned} f_{S_n}(t) &= \sum_{j=n}^{\infty} \left[\left(\frac{\lambda j (\lambda t)^{j-1}}{j!} \right) e^{-\lambda t} + \frac{(\lambda t)^j}{j!} (-\lambda) e^{-\lambda t} \right] \\ &= \lambda e^{-\lambda t} \sum_{j=n}^{\infty} \left[\frac{(\lambda t)^{j-1}}{(j-1)!} - \frac{(\lambda t)^j}{j!} \right] \\ &= \frac{\lambda e^{-\lambda t} (\lambda t)^{n-1}}{(n-1)!} \quad \text{and DONE} \end{aligned}$$

OR using another way:

$$\begin{aligned} f_{S_n}(t)dt &= \mathbf{P}(t \leq S_n \leq t + dt) \\ &= P\left(\underbrace{N(t) = n-1}_{\text{independent}} \text{ and } \underbrace{\text{at least one event in } [t, t+dt]}_{\text{independent}} \right) \\ &= P(N(t) = n-1) P(N(dt) \geq 1) \\ &= \frac{(\lambda t)^{nt} e^{-\lambda t}}{(n-1)!} [\lambda dt + o(dt) + o(dt)] \end{aligned}$$

Dividing the last expression by dt and taking $dt \to 0$,

$$f_{S_n} = \frac{(\lambda t)^{nt} e^{-\lambda t}}{(n-1)!} \left(\lambda + 2\frac{o(dt)}{dt} \right) \xrightarrow{dt \to 0} \frac{(\lambda t)^{nt} e^{-\lambda t}}{(n-1)!}$$

□

The plan is to finish the proof of Proposition 3.4 by calculating the joint density of X_i's. To do so we need the joint density of the S_i's. To this end we will introduce the concept of *order statistic*.

3.2.1. Order Statistic. Let $Y_1 \dots Y_n$ be n random variables. We say that $Y_{(1)} \dots Y_{(n)}$ are the order statistics corresponding to $Y_1 \dots Y_n$ if $Y_{(k)}$ is the k-th smallest value among $Y_1 \dots Y_n$

Lemma 3.5. *If Y_i's are continuous random variables with p.d.f. f then the joint density of the order statistics $Y_{(1)} \dots Y_{(n)}$ is given by*

$$f(Y_1 \dots Y_n) = n! \prod_{i=1}^{n} f(Y_{(i)})$$

PROOF. exercise. See page 66 of the handed notes. □

Theorem 3.6. *The joint density of $(S_1 \dots S_n) = (X_1, X_1 + X_2 \dots \sum_{i=1}^{n} X_i)$ is*

$$f_{S_1 \dots S_n}(t_1 \dots t_n) = \lambda^n e^{-\lambda t_n} I_{\{0 \le t_1 < t_2 \dots < t_n\}}$$

PROOF. Let $0 \le t_1 < t_2 \dots < t_n$, and $\delta > 0$ small enough[2] such that $0 \le t_1 < t_1 + \delta < t_2 < t_2 + \delta \dots < t_n$. Let

$$I_j = (t_j, t_j + \delta)$$

Goal: Find $\mathbf{P}(S_1 \in I_1, S_2 \in I_2 \dots S_n \in I_n)$, then take $\delta \to 0$ to obtain the joint density. Note that terms can be express as follows

$$\begin{cases} S_1 \in I_1 & \text{no event in } (0, t_1) \text{ and 1 event in } I_1 \\ S_2 \in I_2 & \text{no event in } (t_1 + \delta, t_2) \text{ and 1 event in } I_2 \\ \quad \vdots & \\ S_n \in I_n & \text{no event in } (t_{n-1} + \delta, t_n) \text{ and at least 1 event in } I_n \end{cases}$$

[2] In other words, chose δ such that we create non-overlapping intervals.

Then

$$\begin{aligned}
&\mathbf{P}(S_1 \in I_1, S_2 \in I_2 \dots S_n \in I_n) \\
&= \underbrace{e^{-\lambda t}}_{0\in(0,t_1)} \underbrace{\left(\frac{\lambda\delta}{1!}e^{-\lambda(t_2-t_1-\delta)}\right)}_{1\in I_1} \underbrace{e^{-\lambda t}}_{0\in(t_1+\delta,t_2)} \underbrace{\left(\frac{\lambda\delta}{1!}e^{-\lambda(t_2-t_1-\delta)}\right)}_{1\in I_2} \dots e^{-\lambda(t_n-t_{n-1}-\delta)} \underbrace{\left(1-e^{-\lambda\delta}\right)}_{\text{at least 1 in } I_n} \\
&= (\lambda\delta)^{n-1}e^{-\lambda\delta(n-1)}(1-e^{-\lambda\delta})e^{-\lambda t_n}e^{\lambda(n-1)\delta} \\
&= (\lambda\delta)^{n-1}e^{-\lambda t_n}(1-e^{-\lambda\delta})
\end{aligned}$$

Divide[3] by δ^n

$$\begin{aligned}
\frac{\mathbf{P}(S_1 \in I_1, S_2 \in I_2 \dots S_n \in I_n)}{\delta} &= e^{-\lambda t_n}\lambda^{n-1}\underbrace{\frac{1-e^{-\lambda\delta}}{\delta}}_{\to\lambda} \\
&\to \lambda^n e^{-\lambda t_n}
\end{aligned}$$

□

3.2.2. Finishing the proof.

Proof of Proposition 3.4. Note that $X_1 = S_1, X_2 = S_2 - S_1, \dots, X_n = S_n - S_{n-1}$. Therefore we can obtain their distribution from:

$$f_{X_1\dots X_n}(x_1,\dots,x_n) = f_{S_1\dots S_n}\left(x_1, x_1+x_2, \dots, \sum_{i=1}^n x_i\right)|J|\mathbf{1}_{\{0\le x_1\le x_1+x_2\dots\le\sum_{i=1}^n x_i\}}$$

The determinant of the Jacobian J of the transformation is

$$\begin{vmatrix} 1 & 1 & 1\dots & 1 \\ 0 & 1 & 1\dots & 1 \\ & \vdots & \vdots & 1 \\ 0 & 0 & 0\dots & 1 \end{vmatrix} = 1$$

Hence

$$\begin{aligned}
f_{X_1\dots X_n}(x_1\dots x_n) &= f_{S_1,\dots,S_n}\left(x_1, x_1+x_2, \dots, \sum_{i=1}^n x_i\right)\mathbf{1}_{\{0\le x_1\le x_1+x_2\dots\le\sum_{i=1}^n x_i\}} \\
&= \lambda^n e^{-\lambda(x_1+x_2\dots x_n)}\prod_{i=1}^n \mathbf{1}_{\{x_i\ge 0\}} \\
&= \prod_{i=1}^n \lambda e^{-\lambda x_i}\mathbf{1}_{\{x_i\ge 0\}}
\end{aligned}$$

which is the product of n independent exponential distributions. □

[3] Note that $\frac{1-e^{-a}}{a} \xrightarrow{a\to 0} = 1$

Corollary 3.7. Given $S_n = t_n$ the other $n-1$ arrival times $S_1, S_2 \dots S_{n-1}$ have the same distribution as the order statics corresponding to $(n-1)$ independent uniform random variables on the interval $(0, t_n)$.

PROOF.

$$\begin{aligned} f_{S_1 \dots S_{n-1}|S_n}(t_1 \dots t_{n-1}|t_n) &= \frac{f_{S_1 \dots S_n}(t_1 \dots t_n)}{f_{S_n}(t_n)} \\ &= \frac{\lambda^n e^{-\lambda t_n} \mathbf{1}_{\{0 \le t_1 < t_2 \dots < t_n < t\}}}{\frac{\lambda e^{-\lambda t_n} (\lambda t_n)^{n-1}}{(n-1!)}} \\ &= \frac{(n-1)!}{t_n^{n-1}} \mathbf{1}_{\{0 \le t_1 < t_2 \dots < t_n < t\}} \end{aligned}$$

□

Corollary 3.8. Given $N(t) = n$ the n arrival times $S_1, S_2 \dots S_n$ have the same distribution as the order statics corresponding to n independent uniform random variables on the interval $(0, t)$, i.e.

$$f_{S_1 \dots S_n|N(t)}(t_1 \dots t_n|n) = \frac{n!}{t^n} \mathbf{1}_{\{0 \le t_1 < t_2 \dots < t_n < t\}}$$

PROOF. Exercise. □

Proposition 3.9. *Assume that each event of a Poisson(λ) process can be classified as either Type I or Type II event. Furthermore, suppose that if an event occurs at time s then it is classified as being Type I with probability $p(s)$ and Type II with probability $1 - p(s)$.*

If $N_i(t)$ is the number of events of Type i, $i \in \{I,II\}$ by time t, then $N_1(t)$ and $N_2(t)$ are independent Poisson random variables with means(rates) λtp and $\lambda t(1-p)$ respectively, where

$$p = \frac{1}{t} \int_0^t p(s) ds$$

PROOF. Omitted. □

Corollary 3.10. In general if $N(t)$ is poisson(λ) process and events can be categorized into some category type A independently of the original process, then if $N_A(t)$ is the number of events of type A by time t, then $N_A(t)$ is Poisson with rate $\lambda \cdot \int_0^t p_A(s) ds$, where $p_A(s)$ is the probability that one event occurring at time s is of type A.

Note that the original Poisson(λ) process has the mean $\mathbb{E}[N(t)] = \lambda t$.

For the process counting the events of type A, the rate (mean) can be written[4] as $\mathbb{E}[N_A(t)] = \underbrace{\lambda t}_{} \cdot \underbrace{\frac{1}{t}\int_0^t p_A(s)ds}_{}$.

3.3. General Poisson Processes

Definition 3.11. Let $\mathscr{X}$ be a set, $\mathscr{G}$ be a σ-field on $\mathscr{X}$. A counting process on $(\mathscr{X}, \mathscr{G})$ is a stochastic process $\{N(A)\}_{A\in\mathscr{G}}$ with the following properties:

(i)$N(A) \in \{0, 1, 2 \ldots\}$

(ii)$N(\bigcup_{i=1}^{\infty} A_i, \omega) = \sum_{i=1}^{\infty} N(A_i, \omega), \forall w \in \Omega$ and $A_1, A_2 \ldots$ disjoint sets in $\mathscr{G}$

Definition 3.12 (General Poisson Process)**.** Let $(\mathscr{X}, \mathscr{G}, \mu)$ be a measure space, A Poisson process on $(\mathscr{X}, \mathscr{G})$ with intensity μ is a counting process $\{N(A)\}_{A\in\mathscr{G}}$ with

(1) $N(A)$ is a poisson random variable with mean $\mu(A)$
(2) Independent increments, i.e., if $A_1, A_2 \ldots A_n$ are disjoint $\mathscr{G}$ sets in $\mathscr{X}$, then $N(A_1), N(A_2), \ldots N(A_n)$ are independent random variables.

Theorem 3.13. *Let $\{N(A)\}_{A\in\mathscr{G}}$ be a counting process on $(\mathscr{X}, \mathscr{G})$. Let $\mu(A) = \mathbb{E}[N(A)]$ for $A \in \mathscr{G}$. If:*

(1) *$N(\cdot)$ has independent increments (as before),*
(2) *$\forall \epsilon > 0$, there exists $\delta > 0$ such that $\forall$ A with $\mathbb{E}[N(A)] < \delta$,*

$$\frac{\mathbf{P}(N(A) \geq 2)}{\mu(A)} < \epsilon$$

(3) *If $x \in \mathscr{X}$ with $\mu(\{x\}) > 0$ then $N(\{x\}) \sim Poisson(\mu(\{x\}))$*

Then $\{N(A)\}_{A\in\mathscr{G}}$ is a Poisson variable with mean $\mu(A)$

Consequence: If $\{N(A)\}_{A\in\mathscr{G}}$ satisfies the above then

$$P(N(A) = k) = \frac{\mu(A)^k}{k!} e^{-\mu(A)}$$

Example 3.14 (Non-homogenous Poisson Process)**.** This is a simple generalization of the regular Poisson process. The rate is a function of time $\lambda(t)$ instead of λt. In terms of the previous definition $\mathscr{X} = [0, \infty)$, $\mathscr{G} = \mathscr{B}([0, \infty))$ and $\mu(A) = \int_A \lambda(t)dt$. Notice that this process does not have stationary increments anymore.

[4]Note that the expectation is the product between the rate of the original Poisson(λ) process and the probability that the event is of type A.

Example 3.15 (Compound Poisson Process)**.** For each hit i for a Poisson(μ_1) process N on $(\mathscr{X}, \mathscr{G})$ attach the random variables Y_i iid with c.d.f. F, which give rise to the probability measure μ_2[5]. Then the process:

$$Z(A) = \sum_{i=1}^{N(A)} Y_i$$

is a process called the compound poisson process on $[0, \infty) \times \mathbb{R}$ with intensity $\mu = \mu_1 \times \mu_2$

This is the most general definition of the compound process. In the particular case when N is a regular Poisson process we obtain:

$$Z(t) = \sum_{i=1}^{N(t)} Y_i,$$

called the (simple) compound Poisson process.

Proposition 3.16. *If λ is the rate for the Poisson process $N(t)$ and the variables Y_i have mean μ and variance ν^2 then:*

$$\mathbb{E}[Z(t)] = \lambda\mu t, \quad V[Z(t)] = \lambda(\nu^2 + \mu^2)t$$

As an example of occurrence of such a process imagine claims arriving at a health insurance agency, with the time of events modeled by the Poisson process, and with the amount of the claim given by the variables Y_i.

Example 3.17. Consider a system with possible states $\{1, 2 \ldots\}$. Individuals enters the "system" according to a Poisson(λ) process. At any time after the entry, any individual is in some state $i \in \mathbb{N}^*$ ($\mathbb{N}^* = \mathbb{N} \setminus \{0\}$).

Let $\alpha_i(s) = \mathbf{P}\{\text{An individual is in state } i \text{ at time } s \text{ after entry}\}$.

Let $N_i(t)$ be the number of individual in state i at time t. Find $\mathbb{E}[N_i(t)]$.

Solution: We can represent the state of each point of this process as the pair:

$$\left(\underbrace{\text{entry time}}_{\text{poisson process}}, \underbrace{\text{state at time t}}_{\text{the r.v., } Y_i} \right) \in [0, t] \times \mathbb{N}^*.$$

[5]One can obtain the measure from c.d.f. remembering that the Borel sets are generated by intervals and using the relation $\mu((a, b]) = F(b) - F(a)$

An event is of type i if at time t it is in state i. The process N that counts the number of particles on this set $\mathscr{X} = [0,t] \times \mathbb{N}^*$ is a general Poisson process.

Using this definition $N_i(t) = N([0,t] \times \{i\})$. Recalling the theorem 3.13, we have that $N_i(t)$ is a Poisson random variable with mean $\mu([0,t] \times \{i\})$. Therefore, the mean is:

$$\mu([0,t]\times\{i\}) = \lambda \int_0^t p(\text{event at time } s \text{ is of type } i)ds = \lambda \int_0^t \alpha_i(t-s)ds.$$

Let us look at this further. We have, using $r = t - s$:

$$\begin{aligned}
\lambda \int_0^t \alpha_i(t-s) &= \lambda \int_0^t \alpha_i(r)dr \\
&= \lambda \int_0^t \mathbf{P}(\text{invidual is in state } i, \, r \text{ units after its entry})dr \\
&= \lambda \int_0^t \mathbb{E}\left[\mathbf{1}_{\{\text{individual is in state } i, \, r \text{ units after its entry}\}}\right] dr \\
\text{Fubini} \rightarrow &= \lambda \mathbb{E}\left[\underbrace{\int_0^t \mathbf{1}_{\{\ldots\}} dr}_{\text{time spent in state } i \text{ during } [0,t]}\right] \\
&= \lambda \mathbb{E}\left[\text{time spent in state } i \text{ during the first } t \text{ time units}\right]
\end{aligned}$$

Question: What happens as $t \to \infty$? □

Example 3.18 (text 2.22)**.** Cars enter a highway (one way highway) according to a poisson(λ) process in time. Each car has velocity $v(i)$ iid with c.d.f.=F.
Q: Assuming that each car travels at constant velocity, find the distribution of the number of cars on the highway between points a and b (spatial points) at time t?

Solution: We have the entry time and velocity, i.e.,

$$(\text{entry time, velocity}) = (S(i), v(i)) \in [0,\infty) \times [0,\infty)$$

A sample outcome is presented in Figure 1. The position of the car i at time t is $v(i)(t - S(i))$

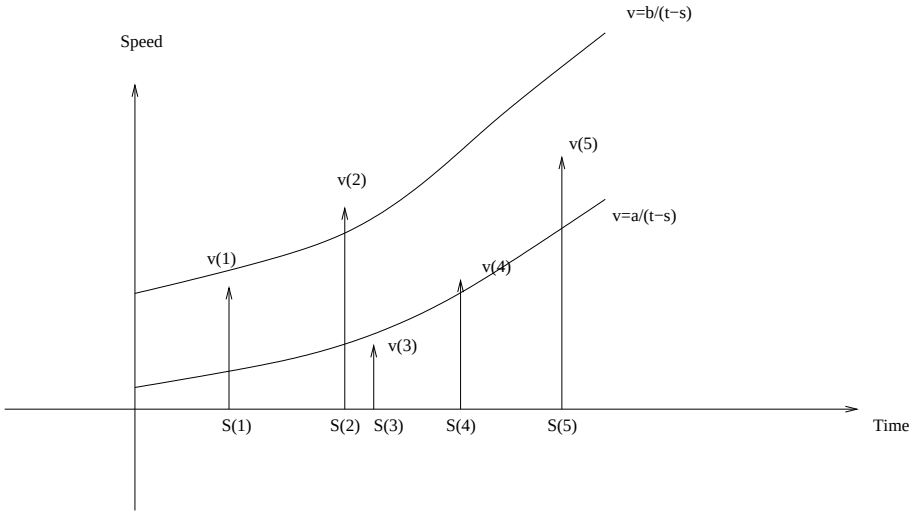

FIGURE 1. Cars enter at $S(i)$ with velocity $v(i)$

We call the event i an event of Type AB if a car entering at s_i with velocity v_i is in $[a, b]$ at time t. Then using Corollary 3.10:

$N(t) =$ number of cars in $[a, b] =$ number of events of type AB

$$N(t) \sim \text{Poisson with the rate} = \lambda \int_0^t p(\text{events enter at } s \text{ is of type } AB)ds$$

What is the probability that a car that arrives at s will be in the interval $[a, b]$ at time t?

$$\begin{aligned} P(\{v : a < v \cdot (t-s) < b\}) &= P\left\{v : \frac{a}{t-s} < v < \frac{b}{t-s}\right\} \\ &= \left[F\left(\frac{b}{t-s}\right) - F\left(\frac{a}{t-s}\right)\right] \mathbf{1}_{[0,t]}(s) \end{aligned}$$

Therefore, $N(t)$ is a Poisson random variable with mean

$$\lambda \int_0^t \left[F\left(\frac{b}{t-s}\right) - F\left(\frac{a}{t-s}\right)\right] ds.$$

Question: What happens as $t \to \infty$? □

3.4. Simulation Techniques. Constructing the Poisson Process.

There are two ways to construct a 1-dim Poisson process

Simplest way : Let $X_1, X_2 \ldots$ iid, exponential(λ) with mean $\frac{1}{\lambda}$. Use X_i as the time between events $i-1$ and i. (Done!)

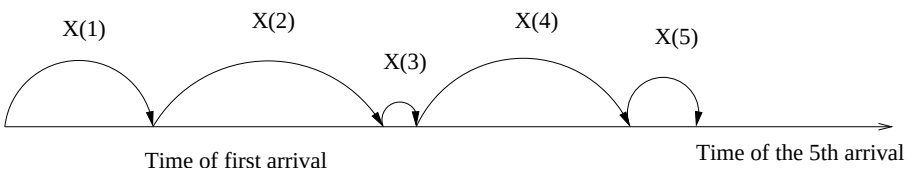

FIGURE 2. Illustration of the construction idea

Generate interval wise: For each time interval $[0,1), [1,2) \dots [t-1,t) \dots$

(1) Simulate[6] $N_I = N([k-1,k)) =$ number of events in $I = [k-1,k)$, this generates the number of events in each interval.
(2) To get the actual times of the events, use the uniform distribution to generate times in each interval. For example if say you obtained $N([0,1)) = 2$ just generate 2 Uniform$(0,1)$ random variables, they are your 2 event times.

As anything in life, the simple way is simple but only works with the 1-dim process. The interval-wise way on the other hand is more complicated but it can be extended to the general Poisson process. The way to do it is straight forward. Suppose we have $(\mathscr{X}, \mathscr{G})$ a measurable space, and μ a σ-finite measure (see Definition 1.11). Partition $\mathscr{X}$ into $\{B_i\}_{i=1}^{\infty}$ such that $\mu(B_i) < \infty$. Then for each B_i get:

(1) $N(B_i) =$ the number of events in B_i which is distributed as a Poisson$(\mu(B_i))$ random variable,
(2) $X_1^{(i)}, X_2^{(i)} \dots$ iid[7] with probability distribution

$$P(X_k^{(i)} \in A) = \mu(A|B_i) = \frac{\mu(A \cap B_i)}{\mu(B_i)}$$

Then for every $A \in \mathscr{G}$ let

$$N(A) = \sum_{i=1}^{\infty} N(A \cap B_i) = \sum_{i=1}^{\infty} \left[\sum_{k=1}^{\infty} \mathbf{1}_{\{X_k^{(i)} \in A \text{ and } N(B_i) \geq k\}} \right] \tag{3.3}$$

Theorem 3.19. *The construction above and* (3.3) *yields a Poisson process with intensity* μ *on* $(\mathscr{X}, \mathscr{G})$.

SKETCH OF THE PROOF. We omit the detailed proof, but we give bellow the important ideas of the proof.

[6] Note that for each interval, $N([k-1,k))$ are iid Poisson$(\lambda \cdot 1)$ random variables.
[7] random points positions in B_i

The countable additivity is satisfied automatically, by the definition of measure. The proof continues demonstrating the following facts:

(1) $N(A) \sim \text{Poisson}(\mu(A))$ for any $A \in \mathscr{G}$
(2) $N(A)$ and $N(B)$ are independent if $A \cap B = \emptyset$

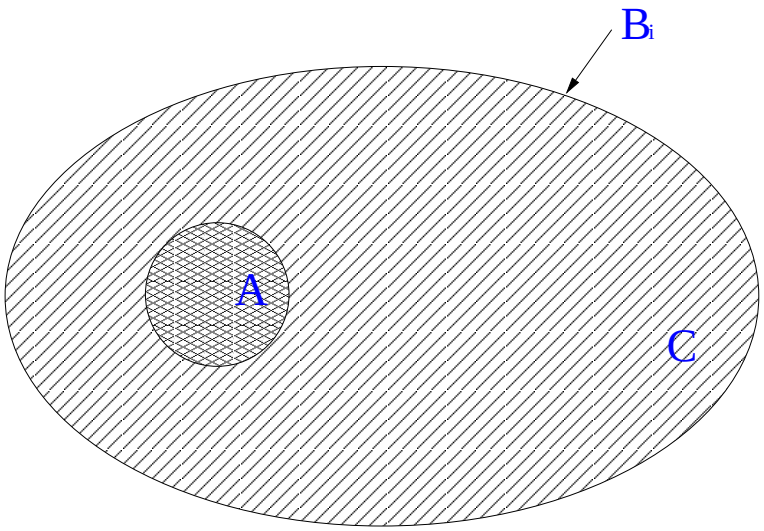

FIGURE 3. Illustration of set A and B_i

First show these properties inside each B_i. To show (2) inside B_i we may proceed as follows. Let $A \subset B_i$ and $C = B_i \backslash A$ (see Figure 3). For integers $a, c \in \mathbb{N}$ we have:

$$
\begin{aligned}
\mathbf{P}(N(A)=a,&N(C)=c) = \mathbf{P}(N(A)=a, N(C)=c, N(B_i)=a+c) \\
&= \mathbf{P}(N(A)=a, N(C)=c | N(B_i)=a+c)\mathbf{P}(N(B_i)=a+c) \\
&= P(N(A)=a|N(B_i)=a+c)\cdot P(N(B_i)=a+c) \qquad (3.4)\\
&= P(N(A)=a|N(B_i)=a+c)\cdot \frac{[\mu(B_i)]^{a+c}}{(a+c)!}e^{-\mu(B_i)} \qquad (3.5)\\
&= \binom{a+c}{a}[\mu(A|B_i)]^a[\mu(C|B_i)]^c\cdot \frac{[\mu(B_i)]^{a+c}}{(a+c)!}e^{-\mu(B_i)} \qquad (3.6)\\
&= \binom{a+c}{a}\frac{[\mu(A\cap B_i)]^a[\mu(c\cap B_i)]^c}{(a+c)!}e^{-\mu(B_i)} \qquad (3.7)\\
&= \frac{[\mu(A\cap B_i)]^a[\mu(c\cap B_i)]^c}{a!c!}e^{-\mu(A\cap B_i)-\mu(C\cap B_i)} \\
&= \underbrace{\frac{[\mu(A\cap B_i)]^a}{a!}e^{-\mu(A\cap B_i)}}_{\text{poisson in } A}\cdot\underbrace{\frac{[\mu(C\cap B_i)]^c}{c!}e^{-\mu(C\cap B_i)}}_{\text{poisson in } C} \\
&= \frac{[\mu(A)]^a}{a!}e^{-\mu(A)}\cdot\frac{[\mu(C)]^c}{c!}e^{-\mu(C)} \\
&= P(N(A)=a)\cdot P(N(C)=c)
\end{aligned}
$$

In (3.4), we removed the redundant information.
In (3.5), we used the Poisson distribution to write the probability for $P(N(B_i)=a+c)$
In (3.6) we used the binomial distribution with $n=a+c$ and $p=\mu(A|B_i)$
In (3.7) by the definition of conditional probability $[\mu(A|B_i)]^a = \frac{[\mu(A\cap B_i]^a}{[\mu(B_i]^a}$

Therefore, $N(A)$ and $N(C)$ are independent. □

Example 3.20 (Astronomy)**.** Consider stars distributed in space according to a 3D Poisson process with intensity, $\lambda\mu$, where μ is the Lebesgue measure[8] on $\mathbb{R}^3$, $\lambda>0$. Let x, y be 3-dim vectors (position).

[8]The Lebesgue measure is the standard way of assigning a length, area or volume to subset of Euclidean space. It is used throughout real analysis, in particular to define Lebesgue integration. Sets which can be assigned a volume are called Lebesgue measurable; the volume or measure of the Lebesgue measurable set A is denoted by $\lambda(A)$. A Lebesgue measure of ∞ is possible, but even so, assuming the axiom of choice, not all subset in $\mathbb{R}^n$ are Lebesgue measurable. The "strange" behavior of non-measurable sets gives rise to such statements as the Banach-Tarski paradox, a consequence of the axiom of choice.

Assume that light intensity exerted at x by a star located at y is

$$f(x, y, \alpha) = \frac{\alpha}{\|x - y\|^2} = \frac{\alpha}{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2},$$

where α is a random parameter depending on the size of the star at y.

Assume that α's associated with stars are iid with common mean μ_α and common variance σ_α^2. Also assume that the combined intensity at x accumulates additively.

Let $Z(x, A)$ be the total intensity at x due to stars in the region A. Then:

$$Z(x, A) = \sum_{i=1}^{N(A)} \frac{\alpha_i}{\|X - Y_i\|^2} = \sum_{i=1}^{N(A)} f(x, y_i, \alpha_i),$$

where $N(A)$ is the number of stars in the region A in space. Note that Y and α are random variables.

We have:

$$\mathbb{E}[Z(x, A)] = \mathbb{E}[N(A)]\mathbb{E}[f(x, Y, \alpha)]. \tag{3.8}$$

We do not prove this result here but note that the expression is a direct consequence of the Wald's equation.

We have that $\mathbb{E}[N(A)] = \lambda\mu(A)$, where $\mu(A)$ is the volume of A.

$$\mathbb{E}[f(x, Y, \alpha)] = \mathbb{E}\left[\frac{\alpha}{||x - Y||^2}\right] = \mathbb{E}[\alpha]\mathbb{E}\left[\frac{1}{||x - Y||^2}\right]$$

Since α and Y are independent. As a consequence of the Poisson Process in space, Y is going to be uniform in A or: $\mathbb{E}[||x - Y||^{-2}] = \frac{1}{\mu(A)} \int_A \frac{1}{||x-y||^2} dy$, then applying the equation (3.8) we have:

$$\begin{aligned}\mathbb{E}[Z(x, A)] &= \lambda\mu(A)\mu_\alpha \frac{1}{\mu(A)} \int_A \frac{1}{||x - y||^2} dy \\ &= \lambda\mu_\alpha \int_A \frac{1}{||x - y||^2} dy\end{aligned}$$

CHAPTER 4

Renewal Processes

In the past I have concentrated on these processes because of the strength of the theorems one can prove. I will state here the majority of the results but we are going to cover a lot less than what this chapter contains.

Example 4.1 (Typical example where renewal process appears)**.** A light bulb in a room keeps burning out. Assume that a mechanism instantaneously replaces the bulb with another one as soon as it burns. Describe the Number of light bulbs replaced by time t.

Let $X_1, X_2 \dots$ iid with c.d.f. F, X_i positive, (non identical zero), with $\mathbb{E}[X_1] = \mu \in (0, \infty]$. These variables will describe the lifetimes of the light bulbs. Define:

$$S_n = \text{time to replace the } n\text{-th bulb}$$

$$S_n = \sum_{i=1}^{n} X_i \sim \underbrace{F * F \dots F}_{n \text{ times}} = F_n$$

Note that F_n means F convoluted[1] itself n times.
We define the renewal process, $N(t)$ as:

$$\begin{aligned} N(t) &= \sup\{n : S_n \leq t\} \\ &= \text{number of renewals up to time } t \end{aligned}$$

Note that a Poisson(λ) process is a renewal process. In that special case the X_i's are exponentially distributed. For a general renewal process, X_i's could have any distribution.
We have the property: $\{N(t) \geq n\} \Leftrightarrow \{S_n \leq t\}$. This is the same result we had for the Poisson(λ) process. Therefore, we can write:

$$\begin{aligned} P(N(t) = n) &= P(N(t) \geq n) - P(N(t) \geq n+1) \\ &= P(S_n \leq t) - P(S_{n+1} \leq t) = F_n(t) - F_{n+1}(t) \end{aligned}$$

[1]Recall $X, Y \sim F, G$ and with pdf f, g then: $X + Y \sim F * G(z) = G * F(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx$

The renewal function The renewal function is the main topic of our study.

$$m(t) = \mathbb{E}[N(t)] = \text{The expected number of renewals by time } t$$

$$\begin{aligned} m(t) = \mathbb{E}[N(t)] &= \mathbb{E}\left[\sum_{i=1}^{\infty} I_{\{S_i \leq t\}}\right] \\ &= \sum_{i=1}^{\infty} \mathbb{E}\left[I_{\{S_i \leq t\}}\right] \\ &= \sum_{i=1}^{\infty} P(S_i \leq t) \\ &= \sum_{i=1}^{\infty} F_i(t) \end{aligned}$$

Thus we just showed that:

$$m(t) = \sum_{i=1}^{\infty} F_i(t) \tag{4.1}$$

Proposition 4.2. *$m(t) < \infty$, for all $0 < t < \infty$ fixed*

Proof. Assume $P(X_k \geq 1) = p > 0$. We will make this assumption. Since $P(X_k = 0) < 1$ then it must exist a positive value α such that $P(X_k \geq \alpha) = p > 0$. If the proof works with $\alpha = 1$ we can later substitute α and the proof will not change significantly.

Let $j - 1 \leq t \leq j$

Claim: $N(t)$ the number of renewals by time $t \leq$ sum of j independent "total" number of trials Geometric(p) random variable. Let us prove the claim. For each bulb k,

If $X_k < 1 =$ throw away the bulb (it counts as a renewal)

If $X_k \geq 1 =$ use the bulb for 1 unit of time then throw it away

If $N^*(t)$ is the number of bulbs replaced by time t using the protocol described above, we obviously have $N^*(t) \geq N(t)$. This proves the claim since $N^*(t)$ has the desired probability distribution.

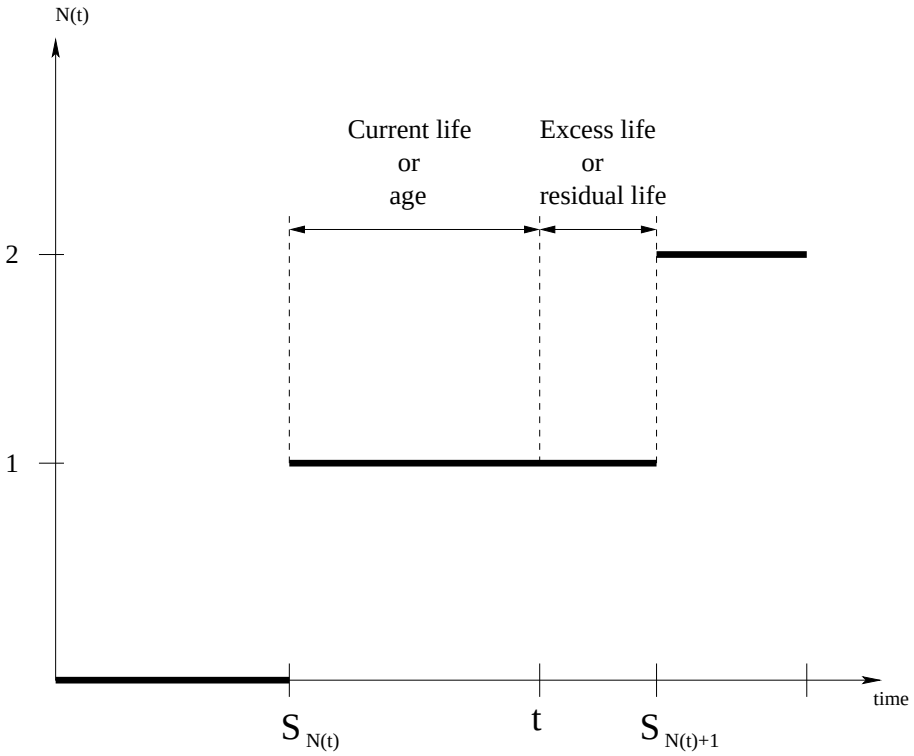

FIGURE 1. Relationship between $S_{N(t)}$, t, and $S_{N(t)+1}$

Therefore, using the claim:

$$\begin{aligned} m(t) = \mathbb{E}[N(t)] &\leq \mathbb{E}[N^*(t)] \\ &= \mathbb{E}[Y_1 + Y_2 \dots Y_j] \\ &= \underbrace{\frac{1}{p} + \frac{1}{p} \dots \frac{1}{p}}_{j \text{ times}} \\ &= \frac{j}{p} < \frac{t+1}{p} < \infty \quad (\text{because } j - 1 < t < j) \end{aligned}$$

We also have:

$$\begin{aligned}\mathbb{E}[N(t)^2] &\leq \mathbb{E}[N^*(t)^2] \\ &\leq \mathbb{E}[(Y_1 + Y_2 \dots Y_j)^2] \\ &= \underbrace{\mathrm{Var}(Y_1 + Y_2 \dots Y_j)}_{\text{negative binomial}} + \underbrace{(\mathbb{E}[Y_1 + Y_2 \dots Y_j])^2}_{\text{known}} \\ &= \frac{j(j-p)}{p^2} + \left(\frac{t+1}{p}\right)^2 < c(t+1)^2 < \infty \quad \text{as well.} \end{aligned} \tag{4.2}$$

□

4.1. Limit Theorems for the renewal process

We will consider limiting results (as $t \to \infty$) for the processes defined thus far.

Proposition 4.3 (Strong Law of Large Numbers for renewal processes)**.** *Using the notation defined earlier,*

$$\frac{N(t)}{t} \to \frac{1}{\mu} \quad a.s.\ as\ t \to \infty$$

Proof of SLLN:. $\frac{N(t)}{t} \to \frac{1}{\mu}$ a.s.
Recall that $S_n = \sum_{i=1}^n X_i$. Then the regular SLLN for random variables gives that $S_n \Rightarrow \frac{S_n}{n} \to \mu$ a.s.

By the definition of $N(t)$ we have $S_{N(t)} \leq t < S_{N(t)+1}$

Divide both sides by $N(t)$

$$\underbrace{\frac{S_{N(t)}}{N(t)}}_{\to\mu \text{ by SLLN}} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)} = \underbrace{\frac{S_{N(t)+1}}{N(t)+1}}_{\to\mu \text{ by SLLN}} \cdot \underbrace{\frac{N(t)+1}{N(t)}}_{\to 1 \text{ a.s.}}$$

which implies $\frac{t}{N(t)} \to \mu$ a.s. OR $\frac{N(t)}{t} \to \frac{1}{\mu}$ a.s.

□

Now we want to obtain a convergence result for $m(t)$. Notice that $m(t) = \mathbb{E}[N(t)]$ and we already have a convergence result for $N(t)$ (this SLLN). Can we get a result about $m(t)$ immediately. Not necessarily as the following example shows.

Example 4.4 (a.s. convergence does not necessarily imply L^1- convergence)**.** Assume that $X_n \xrightarrow{\text{a.s.}} 0$. Is is always true that $\mathbb{E}[X_n] \to 0$?

ANSWER: Not necessary, for example let $U \sim \text{Uniform}[0,1]$, and define $X_n = n\mathbf{1}_{\{U<\frac{1}{n}\}}$. Then we have $X_n \xrightarrow{\text{a.s.}} 0$, but

$$\mathbb{E}[X_n] = n \cdot P\left(\mu < \frac{1}{n}\right) = n \cdot \frac{1}{n} = 1 \to 1 \neq 0$$

□

However for our particular case the implication is true. For the result to be true we need to apply either the dominated convergence theorem or the monotone convergence theorem.

Theorem 4.5 (Elementary renewal theorem)**.** *With the earlier notations we have:*

$$\frac{m(t)}{t} \to \frac{1}{\mu} \text{ as } t \to \infty$$

with the convention $\frac{1}{\infty} = 0$

UNIMAGINATIVE PROOF. Recall that we showed in (4.2) that $\mathbb{E}[N(t)^2] \leq c(t+1)^2$. Thus we have:

$$\mathbb{E}\left[\left(\frac{N(t)}{t}\right)^2\right] \leq \frac{c(t+1)^2}{t^2} \leq 2c \quad \text{which is independent of } t$$

Therefore, $\frac{N(t)}{t}$ is uniformly integrable (since it is in L^2). Thus we get the desired result immediately □

4.1.1. Wald's Theorem. Discrete stopping time. We could just leave the Elementary renewal theorem the way it is, after all we have proven it. However, instead we will prove it again using different concepts which we will use latter on.

The first such new concept is the next theorem which is very very general and very, very useful.

Theorem 4.6 (Wald's Theorem/Identity/Equation)**.** *Let* $X_1, X_2 \ldots$, $W_1, W_2 \ldots$ *be 2 sequence of random variables with* X_k *independent of* W_k *for any fixed* k*. If one of the following conditions is true*

(1) *All* X_k*'s and* W_k*'s are* ≥ 0

(2) $\sum_{k=1}^{\infty} \mathbb{E}[W_k X_k] < \infty$

Then

$$\mathbb{E}\left[\sum_{k=1}^{\infty} W_k X_k\right] = \sum_{k=1}^{\infty} \mathbb{E}[W_k]\mathbb{E}[X_k]$$

Proof of Wald's Theorem: **If the first hypothesis is true** we have:

$$\mathbb{E}\left[\sum_{k=1}^{\infty} X_k W_k\right] \underbrace{=}_{\text{positivity}} \sum_{k=1}^{\infty} \mathbb{E}[X_k W_k]$$

$$\underbrace{=}_{\text{independence}} \sum_{k=1}^{\infty} \mathbb{E}[X_k]\mathbb{E}[W_k]$$

If the second hypothesis is true: $\sum_{k=1}^{\infty} \mathbb{E}[X_k W_k] < \infty$.

Let

$$W_k^+ = W_k I_{\{W_k \geq 0\}}$$
$$W_k^- = -W_k I_{\{W_k < 0\}}$$
$$X_k^+ = X_k I_{\{X_k \geq 0\}}$$
$$X_k^- = -X_k I_{\{X_k < 0\}}$$

Note that

$$W_k = W_k^+ - W_k^-$$
$$X_k = X_k^+ - X_k^-$$

We then have the following:

$$\sum_{k=1}^{\infty} W_k X_k = \sum_{k=1}^{\infty} W_k^+ X_k^+ - \sum_{k=1}^{\infty} W_k^+ X_k^- - \sum_{k=1}^{\infty} W_k^- X_k^+ + \sum_{k=1}^{\infty} W_k^- X_k^-$$

All X_k^+, X_k^-, W_k^+ and W_k^- are positive, then from part 1, we have

$$\mathbb{E}[\sum W_k^+ X_k^+] = \sum \mathbb{E}[W_k^+]\mathbb{E}[X_k^+]$$
$$\mathbb{E}[\sum W_k^+ X_k^-] = \sum \mathbb{E}[W_k^+]\mathbb{E}[X_k^-]$$
$$\mathbb{E}[\sum W_k^- X_k^+] = \sum \mathbb{E}[W_k^-]\mathbb{E}[X_k^+]$$
$$\mathbb{E}[\sum W_k^- X_k^-] = \sum \mathbb{E}[W_k^-]\mathbb{E}[X_k^-]$$

Recombining the terms in the expression above will finish the proof. □

Example 4.7. Let $X_1, X_2 \ldots$ iid with $\mathbb{E}[X_i] = \mu$. Define X_k to be the gain at some game if you actually make the k-th bet. Let

$$W_k = \begin{cases} 1 & \text{if you win} \\ 0 & \text{if you loose} \end{cases}$$

then $\sum_{k=1}^{\infty} X_k W_k$ is the total gain from all bets.

Assume that $X_k > 0$, also assume that W_k is determined by previous bets and by $X_1 \dots X_{k-1}$[2] and maybe some $U \sim$ uniform random variable Let $N = \sum_{i=1}^{\infty} W_k$ to be the number of bets you win. Then Wald's theorem says that if $\mathbb{E}[N] < \infty$ we have

$$\mathbb{E}\left[\sum_{k=1}^{\infty} W_k X_k\right] = \sum_{k=1}^{\infty} \mathbb{E}[W_k] \underbrace{\mathbb{E}[X_k]}_{=\mu}$$

$$= \mu \sum_{k=1}^{\infty} \mathbb{E}[W_k] = \mu \mathbb{E}[N]$$

Note: Think about this and explain to yourself why this result is obvious.

The second important concept is the notion of a stopping time defined next.

Definition 4.8 (Discrete Stopping time)**.** Let $X_1, X_2 \dots$ a sequence of independent random variable. $N \in \{0, 1, 2 \dots\}$ is called a stopping time for this $\{X_n\}_n$ sequence if $\{N = n\}$ is independent of $X_{n+1}, X_{n+2} \dots$. Note that this is true if $\{N = n\}$ is determined only by $X_1, X_2, \dots, X_n$ (or $\{N = n\}$ is measurable with respect to the sigma algebra generated by $X_1, X_2, \dots, X_n$)

Corollary 4.9 (A simpler version of Wald's theorem which we will use for the renewal processes)**.** . For $X_1, X_2 \dots$ iid with $\mu = \mathbb{E}[X_i]$ finite and N a stopping time with $\mathbb{E}[N] < \infty$, then

$$\mathbb{E}[\sum_{k=1}^{N} X_k] = \mathbb{E}[X_i]\mathbb{E}[N] = \mu \mathbb{E}[N]$$

Proof. We wish to apply the general Wald. For this purpose notice that we can write: $\sum_{k=1}^{N} X_k = \sum_{k=1}^{\infty} X_k \mathbf{1}_{\{N \geq k\}}$. In order to apply regular Wald we need to show that $\mathbf{1}_{\{N \geq k\}}$ is independent of X_k.

Remark 4.10. N is a stopping time $\Leftrightarrow$ $\{N \leq n\}$ is independent of $\{X_{n+1}, X_{n+2} \dots\}$.

Proof of this Remark: is an exercise. As a hint note that $\{N \leq n\} = \cup_{k=1}^{n} \{N = k\}$. □

Then $\{N \leq n\}$ is independent of $X_{n+1}, X_{n+2} \dots$ by the remark
OR $\{N > n\}$ is independent of $X_{n+1}, X_{n+2} \dots$
OR $\{N > n-1\}$ is independent of $X_n, X_{n+1} \dots$

[2]In other words, it can depend on previous wins but not on the current

which implies that $\{N \geq n\}$ is independent of $X_n, X_{n+1} \ldots \Rightarrow$ we can use Wald.

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{N} X_k\right] &= \mathbb{E}\left[\sum_{k=1}^{\infty} X_k \mathbf{1}_{\{N\geq k\}}\right] \\
&= \sum_{k=1}^{\infty} \underbrace{\mathbb{E}[X_k]}_{=\mu} \mathbb{E}[\mathbf{1}_{\{N\geq k\}}] \\
&= \mu \sum_{k=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{N\geq k\}}] \\
&= \mu \sum_{k=1}^{\infty} P(N \geq k) \\
&= \mu \begin{pmatrix} P(N=1) & +P(N=2) & +P(N=3) & +\ldots \\ & +P(N=2) & +P(N=3) & +\ldots \\ & & +P(N=3) & +\ldots) \end{pmatrix} \\
&= \mu[1 \cdot P(N=1) + 2 \cdot P(N=2) + 3 \cdot P(N=3) + \ldots] \\
&= \mu \mathbb{E}[N]
\end{aligned}$$

□

4.1.2. Back to the renewal processes. Now the idea is to use the Wald's theorem we just prove. To this end we need to find a stopping time for the inter-arrival times. So, the next question comes naturally: "Is $N(t) = \sup\{n : S_n \leq t\}$ a stopping time for $X_1 \ldots X_n \ldots$"

Short answer: No, since there could be an event happen between the n-th event and t.

Mathematical answer:

$$\begin{aligned}
\{N(t) = n\} &= \{S_n \leq t < S_{n+1}\} \\
&= \{X_1 + X_2 \ldots + X_n \leq t < \underbrace{X_1 + X_2 \ldots X_n + X_{n+1}}_{\text{not independent of } X_{n+1}}\}
\end{aligned}$$

Note that the event is determined by X_{n+1}, hence $\{N(t) = n\}$ can't be a stopping time.

New question: "Is $N(t) + 1$ a stopping time?"

Answer: Yes, note that

$$\begin{aligned}
\{N(t) + 1 = n\} &= \{N(t) = n - 1\} = \{S_{n-1} \leq t < S_n\} \\
&= \{X_1 \ldots + X_{n-1} \leq t < X_1 \ldots X_{n-1} + X_n\}
\end{aligned}$$

Since everything inside the last $\{\cdot\}$ does not contain terms of type $n+1$ or larger, $\{N(t) + 1\}$ is a stopping time.

At this point we are in the position to give the alternate proof to the elementary renewal theorem. We will restate the theorem first.

Theorem 4.11 (Elementary Renewal Theorem)**.** *Let* $m(t) = \mathbb{E}[N(t)]$.

$$\frac{m(t)}{t} \xrightarrow{a.s.} \frac{1}{\mu} \quad (= 0 \text{ if } \mu = \infty)$$

Alternate Proof to the Elementary Renewal Theorem. Since $N(t)+1$ is a stopping time, from Wald we have

$$\mathbb{E}[S_{N(t)+1}] = \mathbb{E}[N(t)+1]\mathbb{E}[X] = \mu(m(t)+1) \tag{4.3}$$

Claim 1. $\liminf_{t\to\infty} \frac{m(t)}{t} \geq \frac{1}{\mu}$

Claim 1 proof: By definition, $S_{N(t)+1} > t \Rightarrow \mathbb{E}[S_{N(t)+1}] > t$. Then using (4.3) we obtain:

$$\mu(m(t)+1) > t$$
$$\frac{m(t)}{t} > \frac{1}{\mu} - \frac{1}{t}$$

Take lim inf on both sides

$$\liminf_{t\to\infty} \frac{m(t)}{t} \geq \liminf_{t\to\infty} \left(\frac{1}{\mu} - \frac{1}{t}\right) = \frac{1}{\mu}$$

□

Claim 2. $\limsup \frac{m(t)}{t} \leq \frac{1}{\mu}$

Claim 2 proof: Fix $M > 0$ constant. Let

$$\bar{X}_k = \begin{cases} X_k & \text{if } X_k \leq M \\ M & \text{if } X_k > M \end{cases}$$

Let $\mu_M = \mathbb{E}[\bar{X}_k]$, $\bar{N}(t)$ to be the number of renewals up to t with lifetimes $\bar{X}_k$.

Note that $\bar{N}(t) \geq N(t)$ (due to shorter life times[3])

$$\bar{m}(t) = \mathbb{E}[\bar{N}(t)] \geq m(t)$$

Now look at the Figure 2 on page 62 which represents the behavior of the new process at t. Note that since we bounded the interarrival times by M we have:

$$\bar{S}_{\bar{N}(t)+1} > t$$
$$\bar{S}_{\bar{N}(t)+1} \leq t + M$$

[3]Life span is limited by the upper bound M.

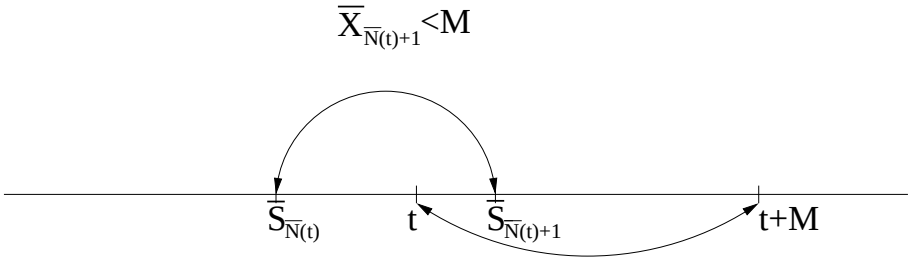

FIGURE 2. Representation of the new process.

If we use (4.3) for $\bar{N}(t)$ we get: $\mathbb{E}[\bar{S}_{\bar{N}(t)+1}] = \mu_M(\bar{m}(t)+1)$. Therefore:

$$\begin{aligned}\mu_M(\bar{m}(t)+1) &\leq t+M\\ \frac{\bar{m}(t)+1}{t} &\leq \frac{1}{\mu_M}+\frac{M}{t\mu_M}\\ \frac{\bar{m}(t)}{t} &\leq \frac{1}{\mu_M}+\frac{M}{t\mu_M}-\frac{1}{t}\end{aligned}$$

Apply lim sup in both sides to get:

$$\limsup_{t\to\infty}\frac{\bar{m}(t)}{t} \leq \frac{1}{\mu_M}$$

which implies

$$\limsup_{t\to\infty}\frac{m(t)}{t} \leq \limsup_{t\to\infty}\frac{\bar{m}(t)}{t} \leq \frac{1}{\mu_M}.$$

But this holds for any $M>0$. Therefore take $M\to\infty$ and using that $\lim_{M\to\infty}\frac{1}{\mu_M}=\frac{1}{\mu}$, we conclude: $\limsup_{t\to\infty}\frac{m(t)}{t}=\frac{1}{\mu}$. □

Now combining the two claims will finish the proof. □

Finally we will give a convergence in distribution result similar with the regular Central Limit Theorem.

Theorem 4.12 (Renewal Central Limit Theorem)**.** *Let $X_1, X_2 \ldots$ i.i.d., positive, $\mu=\mathbb{E}[X_k]$, $\sigma^2 = Var(X_k)<\infty$*

$$S_n=\sum_{i=1}^{n}X_i, \quad N(t)=\sup\{n : S_n\leq t\}$$

Then

$$P\left(\frac{N(t)-\frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}}<y\right)\xrightarrow{t\to\infty}\Phi(y)$$

where $\Phi(y)$ is the c.d.f. of $N(0,1)$

In other words, when t is large $N(t) \sim N(\frac{t}{\mu}, \frac{\sigma^2 t}{\mu^3})$

Proof. Fix y.

$$\begin{aligned} P\left(\frac{N(t)-\frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} < y\right) &= P\left(N(t) < \underbrace{\frac{t}{\mu} + y\sigma\sqrt{\frac{t}{\mu^3}}}_{=r_t}\right) \\ &= P(N(t) < r_t) \\ &= P(S_{r_t} > t) \end{aligned}$$

where it is obvious the definition of r_t. Note here that what we wrote only works if r_t is an integer. If r_t is not an integer, take $\tilde{r}_t = [r_t] + 1$. Then

$$P(N(t) < r_t) = P(N(t) < \tilde{r}_t) = P(S_{\tilde{r}_t} > t)$$

We can use the classic central limit theorem to complete the proof.

$$\begin{aligned} P(S_{\tilde{r}_t} > t) &= P\left(\frac{S_{\tilde{r}_t} - \tilde{r}_t\mu}{\sigma\sqrt{\tilde{r}_t}} > \frac{t - \tilde{r}_t\mu}{\sigma\sqrt{\tilde{r}_t}}\right) \\ &= \lim_{t\to\infty} \Phi\left(\frac{t - \tilde{r}_t\mu}{\sigma\sqrt{\tilde{r}_t}}\right) \end{aligned}$$

The proof of the theorem will end if we show that $\frac{t-\tilde{r}_t\mu}{\sigma\sqrt{\tilde{r}_t}} \to -y$

We have that:

$$\tilde{r}_t = r_t + \underbrace{\{1 - \{r_t\}\}}_{=\Delta_t \in [0,1)}$$

where we used the notation $\{x\}$ for the fractional part of x. This implies:

$$\frac{t - \tilde{r}_t\mu}{\sigma\sqrt{\tilde{r}_t}} = \frac{t - r_t\mu - \Delta_t\mu}{\sigma\sqrt{r_t + \Delta_t}}$$

Recall that $r_t = \frac{t}{\mu} + y\sigma\sqrt{\frac{t}{\mu^3}}$. Therefore, we continue:

$$= \frac{t - \Delta_t\mu - \left(t + y\sigma\sqrt{\frac{t}{\mu}}\right)}{\sigma\sqrt{\Delta_t + \frac{t}{\mu} + \sigma y\sqrt{\frac{t}{\mu^3}}}} \xrightarrow{t\to\infty} \frac{-y\sigma}{\sigma} = -y$$

□

4.2. Discrete Renewal Theory. Blackwell theorem.

We will start this section with a motivating example.

Example 4.13 (Block Replacement Policy)**.** Consider a light-bulb with lifetime X, a random variable. Due to economic reasons, it might be cheaper on a per bulb basis to replace all the bulbs instead of just the one that breaks. A block replace policy does just that by fixing a time period K and replacing bulb's as they failed at times $1, 2 \dots K-1$ and at K replacing everything regardless the condition of the bulb. Let

$$c_1 = \text{replacement cost per bulb (block replacement)}$$
$$c_2 = \text{replacement cost per bulb (failure replacement)}$$

where obviously $c_1 < c_2$. Let $N(n)$ to be the number of replacements up to time n for *1 bulb*, let $m(n) = \mathbb{E}[N(n)]$

For one bulb the expected cost is $c_2 m(k-1) + c_1$. Then, the mean cost per unit of time is:

$$\frac{\text{mean cost}}{\text{unit time}} = \frac{c_2 m(k-1) + c_1}{K}$$

Since the replacements take place only at the beginning of the day we are only interested in discrete variables to describe the lifetime of a lightbulb. Suppose that X has the distribution $P(X = k) = p_k$, $k = 1, 2 \dots$. Fix $n \leq K$. Look at X_1 the lifetime of the first lightbulb. Obviously, if $X_1 > n$ there was no replacement by time n. If $X_1 = k \leq n$ then we will have $m(n-k)$ expected replacements in the later time period. Therefore, we can write conditioning on the lifetime of the first bulb:

$$\begin{aligned} m(n) &= \sum_{k=n+1}^{\infty} p_k \cdot 0 + \sum_{k=1}^{n} p_k [1 + m(n-k)] \\ &= \sum_{k=1}^{n} p_k [1 + m(n-k)] \\ &= F_X(n) + \sum_{k=1}^{n-1} p_k m(n-k), \end{aligned}$$

where $F_X(\cdot)$ is the c.d.f. of X. Then we obtain recursively:

$$\begin{aligned} m(0) &= 0 \\ m(1) &= F_X(1) + p_1 m(0) = p_1 \\ m(2) &= F_X(2) + p_1 m(1) + p_2 m(0) = p_1 + p_2 + p_1^2 \\ &\vdots \quad \text{etc.} \end{aligned}$$

Example 4.14 (continues the example above)**.** Let us look to a numerical example of the problem above. Suppose that X can only take values $\{1, 2, 3, 4\}$ with $p_1 = 0.1$, $p_2 = 0.4$, $p_3 = 0.3$, $p_4 = 0.2$, furthermore the costs are $c_1 = 2$, $c_2 = 3$. Find the optimal replacement policy.

Using the formulas above we can calculate:

$$m(1) = 0.1, \quad m(2) = 0.51, \quad m(3) = 0.891, \quad m(4) = 1.3231$$

Using these numbers we will try to minimize the expect cost,

$$\text{cost} = \frac{c_1 + c_2 m(K-1)}{K} \leftarrow \text{We will try different K's to get the minimum}$$

We will obtain a table of cost as a function of K as:

TABLE 1. default

K	cost
1	2.00
2	1.15
3	1.17
4	1.16
5	1.19

Hence the optimal replacement policy is at $K = 2$. We can also continue the calculation of m's:

$$\begin{aligned} m(5) =& 1.6617, m(6) = 2.0647, m(7) = 2.4463, m(8) = 2.8336, \\ m(9) =& 3.2136, m(10) = 3.6016, \ldots \end{aligned}$$

Now we can calculate u_n the probability that a replacement occurs in period n as:

$$u_n = m(n) - m(n-1).$$

Calculating u_n's for the values given we can see that pretty quickly we have

$$u_n \approx \frac{1}{\mu} = 0.3846.$$

This fact will be explained by the next theorem.

Let us assume that we have a renewal process with non negative integer valued lifetimes, X with $P(X = k) = p_k$, $k = 0, 1, 2 \dots$

Definition 4.15. X an integer random variable is called a *lattice* if there $\exists d \geq 0$ such that $p_k > 0, \forall k$ not a multiple of d. The largest d with the property that $\sum_{n=1}^{\infty} p_{nd} = 1$ is called the period of X. In effect:

$$d = \text{g.c.d.}\{k : p_k > 0\}^{4}$$

If g.c.d.$\{k : p_k > 0\} = 1$ then X is called a non lattice random variable. Also if X, a lattice random variable has c.d.f F, then F is called a lattice.

Example 4.16. Consider the two simple example below

- $p_2 = p_4 = p_6 = \frac{1}{3}$ lattice distribution.
- $p_3 = p_7 = \frac{1}{2}$ non-lattice distribution.

In the previous example we have seen how to establish the equation

$$m(n) = F_X(n) + \sum_{k=1}^{n-1} p_k m(n-k).$$

(Note that if lifetimes are allowed to be zero the equation is a little different.)

However, this equation constitutes a particular example of a *renewal equation* (discrete case). In general a discrete renewal equation looks like:

$$v_n = b_n + \sum_{k=0}^{n} p_k v_{n-k}, \tag{4.4}$$

where v_i's are unknowns and p_i's are probabilities. Note that this form of equation has a unique solution, e.g. $v_0 = \frac{b_0}{1-p_0}$, $v_1 = \frac{b_1 + p_1 v_0}{1-p_0}$, etc.

Let u_n be the expected number of renewals that take place in period n. We have said in the example that $u_n = m(n) - m(n-1)$. This is only true if lifetimes are nonzero and therefore at most one renewal occurs in any 1 time period. This is easy to show:

$$\begin{aligned} u_n &= \mathbf{P}\{\text{One renewal occurred at } n\} \\ &= \mathbb{E}[\mathbf{1}_{\{\text{One renewal occurred at } n\}}] \\ &= \mathbb{E}[N(n) - N(n-1)] = m(n) - m(n-1) \end{aligned}$$

We have seen in the previous example that this u_n got closer and closer to $1/\mu$. The next theorem formalizes this fact and generalizes it.

[4] The greatest common denominator of the set of integers.

Theorem 4.17 (Blackwell renewal theorem)**.** *Using the notations defined thus far we have:*

(1) $u_n \to \frac{1}{\mu}$ *as* $n \to \infty$.
(2) *If* $X_0 \geq 0$ *is a "delay" variable, and* $X_1, X_2, \ldots \geq 0$ *are i.i.d. lifetimes independent of* X_0 *with* $\mathbb{E}X_1 = \mu$ *and non-lattice distribution then:*

$$m(t+a) - m(t) \to \frac{a}{\mu}, \quad \text{as } t \to \infty.$$

Note that $m(t+a) - m(t)$ *is the expected number of renewals in the interval* $[t, t+a]$.
(3) *If* X_i*'s are lattice random variables with period* d*, and* $X_0 = 0$ *then:*

$$\mathbb{E}[\textit{Number of renewals at } nd] \to \frac{d}{n} \quad , n \to \infty$$

Remark 4.18. About the theorem.

(1) Even though the section was started with an example of a discrete renewal process, the part (2) of the Blackwell theorem applies to **any** non-lattice distribution. This includes any continuous distribution.
(2) All the parts of the theorem are true if $\mu = \infty$ $(1/\infty = 0)$.
(3) If $X_i > 0$, part (3) is $\Leftrightarrow \mathbf{P}\{\text{Renewal at } nd\} \to d/\mu$

PROOF. Not proven. □

Write for an infinitesimal increment dy:

$$\begin{aligned} m(dy) &= \left(\underbrace{dm(y)}_{\text{Notation used sometimes}} \right) = m(y+dy) - m(y) \\ &= \mathbb{E}[\text{Number of renewals in the interval } (y, y+dy]] \end{aligned}$$

This is the *renewal measure.* The Blackwell renewal theorem says that:

$$m(dy) \simeq \frac{1}{\mu} dy.$$

Lemma 4.19. *We have:*

$$m(dy) = \sum_{n=0}^{\infty} \mathbf{P}\left(S_n \in (y, y+dy]\right) \tag{4.5}$$

PROOF. The proof is straightforward (here we use a delay, therefore the sum starts from $n = 0$):

$$\begin{aligned} m(dy) &= \mathbb{E}[N(y+dy) - N(y)] = \mathbb{E}[N(y+dy)] - \mathbb{E}[N(y)] = \\ &= \sum_{n=0}^{\infty} \mathbf{P}\left(N(y+dy) \geq n\right) - \sum_{n=0}^{\infty} \mathbf{P}\left(N(y) \geq n\right) \\ &= \sum_{n=0}^{\infty} \mathbf{P}\left(S_n \leq y+dy\right) - \sum_{n=0}^{\infty} \mathbf{P}\left(S_n \leq y\right) \\ &= \sum_{n=0}^{\infty} \mathbf{P}\left(S_n \in (y, y+dy]\right) \end{aligned}$$

□

Many applications of the renewal theorem are concerned with the behavior of the process near a large time t. We need a final key of the puzzle before we proceed with the study of such applications and this key is provided in the next section.

4.3. The Key Renewal Theorem

This is the main result used in applications of the renewal processes. We will start with a definition.

Definition 4.20 (Directly Riemann Integrable function)**.** A function $h : [0, \infty) \to \mathbb{R}$ is called a Directly Riemann Integrable (DRI) function if the upper and lower mesh δ Darboux sums are finite and have the same limit as $\delta \to 0$.

Reminder of lower (and upper) Darboux sum LDS (and UDS):
Let $\pi = (t_0 = 0 < t_1 < t_2 < \ldots)$ be a partition of $[0, \infty)$, with $max_i(t_i - t_{i-1}) \leq \delta$. Define:

$$LDS(h, \pi, \delta) = \sum_{n=1}^{\infty} \inf_{t \in [t_{n-1}, t_n]} h(t)(t_n - t_{n-1})$$

$$UDS(h, \pi, \delta) = \sum_{n=1}^{\infty} \sup_{t \in [t_{n-1}, t_n]} h(t)(t_n - t_{n-1})$$

Example 4.21 (Example of Riemann integrable function which is **not** DRI)**.** Let:

$$h(s) = \sum_{k=1}^{\infty} \mathbf{1}_{\{k \leq s < k + \frac{1}{k^2}\}}$$

Make a plot of this function to see what is happening. We have that:

$$\int_0^\infty h(s)ds = 1 + \frac{1}{4} + \frac{1}{9} + \cdots = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty,$$

so this function is Riemann integrable. However it is not DRI. Take the partition $\pi = (t_0 = 0, t_1 = \delta, t_2 = 2\delta, \ldots, t_n = n\delta, \ldots)$. Then:

$$\begin{aligned} UDS(h, \pi, \delta) &= \sum_{n=1}^{\infty} \sup_{t \in [(n-1)\delta, n\delta]} h(t)(n\delta - (n-1)\delta) \\ &= \delta \sum_{n=1}^{\infty} \sup_{t \in [(n-1)\delta, n\delta]} h(t) \end{aligned}$$

For any δ no matter how small but positive the last term is an infinite sum of 1's which is infinite.

Proposition 4.22. *The following are sufficient conditions for a function to be DRI:*

(1) $h(t) \geq 0$, $\forall t > 0$
(2) h *is nonincreasing*
(3) $\int_0^\infty h(t)dt < \infty$

Proof. Not given. □

Now we are in the postion to be able to state the main theorem of this section.

Theorem 4.23 (The Key Renewal Theorem)**.** *For non-lattice $X_1, X_2, \ldots$ (any X_0 "delay" is fine) and if h is a DRI function we have:*

$$\lim_{t \to \infty} \int_0^t h(t-y)m(dy) = \frac{1}{\mu} \int_0^\infty h(t)dt$$

Proof. Skipped. □

This is a very powerful theorem. We shall see its application in the next section.

4.4. Applications of the Renewal Theorems

Refer back to Figure 1 on page 55. We can see there the current age at time t and the remaining lifetime at t. Applications are concerned with these quantities when t is large. So the question is: can we get distributions for these quantities? For example:

(a) $\mathbf{P}$(Age at time t of the current item $> x) = \mathbf{P}(A(t) > x)$
(b) $\mathbf{P}$(Remaining lifetime of the item in use at $t > x) = \mathbf{P}(Y(t) > x)$

(c) $\mathbf{P}(\text{Total age of the item in use at } t > x) = \mathbf{P}(X_{N(t)+1} > x)$,

where we have use the obvious notations $A(t)$ to denote the age of the item in use at t and $Y(t)$ to denote the residual life for the item in use at t.

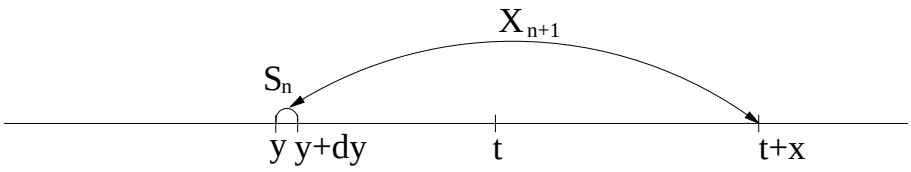

FIGURE 3. Deduction of the formula.

We will look at the process $Y(t)$ for exemplification (see Figure 3 on page 70). Recall that X_0 is the delay and we will use the convention $S_0 = X_0$. Note that this renewal is counted in the renewal process $N(t)$. We have:

$$\begin{aligned}
\mathbf{P}(Y(t) > x) &= \mathbf{P}(S_{N(t)+1} - t > x) = \mathbf{P}(N(t) = 0, X_0 > t + x) \\
&\quad + \sum_{n=1}^{\infty} \int_0^t \mathbf{P}(N(t) = n, S_{n-1} \in (y, y + dy], X_n > t + x - y) \\
&= \mathbf{P}(X_0 > t + x) + \sum_{n=1}^{\infty} \int_0^t \mathbf{P}(S_{n-1} \in (y, y + dy], X_n > t + x - y) \\
&= (1 - F_0(t + x)) + \sum_{n=0}^{\infty} \int_0^t \mathbf{P}(S_n \in (y, y + dy]) \mathbf{P}(X_{n+1} > t + x - y)
\end{aligned}$$

Using the notation:

$$\overline{F}(x) = 1 - F(x)$$

we continue:

$$\begin{aligned}
\mathbf{P}(Y(t) > x) &= \overline{F}_0(t + x) + \int_0^t \overline{F}(t + x - y) \sum_{n=0}^{\infty} \mathbf{P}(S_n \in (y, y + dy]) \\
&= \overline{F}_0(t + x) + \int_0^t \overline{F}(t + x - y) m(dy) \\
&= \overline{F}_0(t + x) + \int_0^t h(t - y) m(dy),
\end{aligned}$$

where we have used the Lemma 4.19 and we used the notation $h(s) = \overline{F}(s + x)$. Using now the fact that $\overline{F}_0(t + x) \xrightarrow{t \to \infty} 0$ (argue this

yourselves) a direct application of the Key Renewal Theorem (KRT) 4.23 will yield:

$$\mathbf{P}(Y(t) > x) \xrightarrow{t\to\infty} \frac{1}{\mu}\int_0^\infty \overline{F}(s+x)ds = \frac{1}{\mu}\int_x^\infty \overline{F}(y)dy$$

This result is significant enough to make it a proposition.

Proposition 4.24. *Let $A(t)$ be the age at t of the item and let $Y(t)$ be the residual life of the item alive at t. Then if F is the c.d.f. of the lifetimes with mean lifetime μ, then the distributions of $A(t)$ and $Y(t)$ for t large have densities proportional with:*

$$f(y) = \frac{\overline{F}(y)}{\mu}$$

PROOF. For $Y(t)$ the result is clear since from above we have:

$$\mathbf{P}(Y(t) \le x) \xrightarrow{t\to\infty} = \int_{-\infty}^x \frac{\overline{F}(y)}{\mu}dy$$

For $A(t)$ note that we have:

$$\{A(t) > x\} \Leftrightarrow \{Y(t-x) > x\} \quad \text{(No renewal in } [t-x, t]),$$

therefore

$$\lim_{t\to\infty} \mathbf{P}\{A(t) > x\} = \lim_{t\to\infty} \mathbf{P}\{Y(t-x) > x\} = \frac{1}{\mu}\int_x^\infty \overline{F}(y)dy$$

□

Remark 4.25. If the distribution of the delay has this special form: $\mathbf{P}(X_0 > x) = \frac{1}{\mu}\int_x^\infty \overline{F}(y)dy$ then $m(t) = \frac{t}{\mu}$ and the process is stationary (meaning that it looks the same regardless when you start observing it).

4.5. Special cases of Renewal Processes. Alternating Renewal process. Renewal Reward process.

4.5.1. The Alternating Renewal process. Let $\{(Z_n, Y_n)\}_{n=1}^\infty$ be i.i.d. pairs of random variables[5]. Note that the pairs for $i \neq j$ are independent but Z_n and Y_n can be dependent.

Let $X_n = Z_n + Y_n$. Let $S_n = \sum_{i=1}^n X_i$ which will give the renewal process.

The story: The Z_i's represent the lightbulb lifetimes or the time that the system is ON, and the Y_i's represent the replacement times or the time that the system is OFF.

[5]Here (Z_1, Y_1) (delay) is allowed to have a different distribution than the rest

Denote the c.d.f of Y_i's with G, the c.d.f of Z_i's with H and the c.d.f of X_i's with F.

Theorem 4.26. *If $\mathbb{E}[X_n] < \infty$ and F is non-lattice we have:*

$$\mathbf{P}(\textit{The system is ON at time } t) \xrightarrow{t\to\infty} \frac{\mathbb{E}(Z_n)}{\mathbb{E}(X_n)} = \frac{\mathbb{E}(Z_n)}{\mathbb{E}(Z_n)+\mathbb{E}(Y_n)}$$

Proof.

$$\begin{aligned}
\mathbf{P}(\text{ON at time } t) &= \mathbf{P}(Z_1 > t) + \sum_{n=0}^{\infty}\int_0^t \mathbf{P}(S_n \in (y, y+dy], Z_{n+1} > t-y) \\
&= \overline{H}_1(t) + \sum_{n=0}^{\infty}\int_0^t \mathbf{P}(S_n \in (y, y+dy])\mathbf{P}(Z_{n+1} > t-y) \\
&= \overline{H}_1(t) + \int_0^t \overline{H}(t-y)\sum_{n=0}^{\infty}\mathbf{P}(S_n \in (y, y+dy]) \\
&= \overline{H}_1(t) + \int_0^t \overline{H}(t-y)m(dy) \\
&\xrightarrow[\text{KRT}]{t\to\infty} \overline{H}_1(\infty) + \frac{1}{\mu}\int_0^\infty \overline{H}(t)dt
\end{aligned}$$

However, $\mathbb{E}[Z] = \int_0^\infty \mathbf{P}(Z > z)dz = \int_0^\infty \overline{H}(z)dz$ and $\mathbb{E}[X] = \mu$ so we are done. □

Example 4.27. We have already seen that the distribution of $A(t)$ has density $\overline{F}(y)/\mu$. We will obtain this distribution again using the previous theorem about alternating renewal processes. Please read the next derivations since they provide examples of using this most useful theorem.

Once again we will deduce $\mathbf{P}(A(t) > x)$. Fix $x > 0$. Say that the system is ON during the first x units of each lifetime and OFF the rest of that time. Mathematically, using the notation of the alternating renewal processes:

$$\begin{aligned}
Z_k &:= X_k \wedge x = \min(X_k, x) \\
Y_k &= X_k - Z_k
\end{aligned}$$

Then the theorem says:

$$\mathbf{P}(\text{System is ON at time t}) = \mathbf{P}(A(t) < x) \to \frac{\mathbb{E}(Z_n)}{\mu}$$

But we can calculate the limit since:

$$\begin{aligned}\mathbb{E}(Z_n) &= \int_0^\infty \mathbf{P}(Z_n > y)dy = \int_0^x \mathbf{P}(Z_n > y)dy + \int_x^\infty \mathbf{P}(Z_n > y)dy \\ &= \int_0^x \mathbf{P}(X_n > y)dy = \int_0^x \overline{F}(y)dy,\end{aligned}$$

which will give the density and finish the solution.

Example 4.28 (Limiting distribution of the current lifetime $X_{N(t)+1}$)**.** We want to calculate $\mathbf{P}(X_{N(t)+1} > x)$. Fix x. Construct an alternating renewal process using:

$$Z_n = X_n \mathbf{1}_{\{X_n > x\}}, \quad Y_n = X_n \mathbf{1}_{\{X_n \leq x\}}$$

Then:

$$\mathbf{P}(\text{System is ON at time t}) = \mathbf{P}(X_{N(t)+1} > x) \to \frac{\mathbb{E}(Z_n)}{\mu}$$

Again we can calculate:

$$\begin{aligned}\mathbb{E}(Z_n) &= \int_0^\infty \mathbf{P}(Z_n > y)dy = \int_0^x \mathbf{P}(Z_n > y)dy + \int_x^\infty \mathbf{P}(Z_n > y)dy \\ &= \int_0^x \mathbf{P}(X_n > x)dy + \int_x^\infty \mathbf{P}(X_n > y)dy \\ &= x\mathbf{P}(X_n > x) + \int_0^\infty \overline{F}(y)dy \\ &= \int_0^\infty y dF(y) \quad \text{(Integrating by parts)}\end{aligned}$$

which will give the limiting distribution:

$$\mathbf{P}(X_{N(t)+1} > x) \to \frac{\int_0^\infty y dF(y)}{\mu}$$

Recall that if we denote $Y(t)$ the excess lifetime, we have already found its limiting distribution:

$$P(Y(t) > x) \to \frac{1}{\mu} \int_x^\infty \bar{F}(t)dy$$

We would like to find its expectation, or the limiting expected excess life, $\mathbb{E}[Y(t)]$. A first guess would be obviously the expectation of the previous distribution:

$$\mathbb{E}[Y(t)] = \int_0^\infty P(Y(t) > x)dx \to \frac{1}{\mu} \int_0^\infty \int_x^\infty \bar{F}(y)dydx$$

The guess turns out to be correct but we need to prove this.

Proposition 4.29. *If X is non-lattice with $\mathbb{E}[X^2] < \infty$, then*

$$\lim_{t\to\infty} \mathbb{E}[Y(t)] = \frac{\mathbb{E}[X^2]}{2\mu}$$

Note that one can show $\frac{\mathbb{E}[X^2]}{2\mu}$ and $\frac{1}{\mu}\int_0^\infty \int_x^\infty \bar{F}(y)dydx$ are the same quantities using a change the order of integration then integrating by parts.

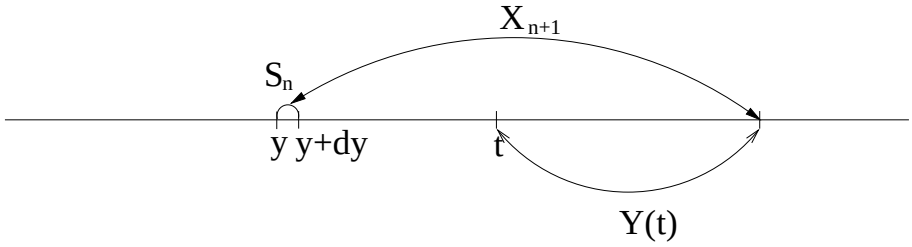

FIGURE 4. Relationship between X_{n+1} and $Y(t)$

PROOF. We can go ahead and calculate:

$$\begin{aligned}
\mathbb{E}[Y(t)] &= \sum_{n=0}^{\infty} \mathbb{E}[Y(t)\mathbf{1}_{\{N(t)=n\}}] \\
&= \mathbb{E}[Y(t)\mathbf{1}_{\{N(t)=0\}}] + \sum_{n=1}^{\infty} \mathbb{E}[Y(t)\mathbf{1}_{\{N(t)=n\}}] \\
&= \mathbb{E}[(X_1 - t)\mathbf{1}_{\{X_1>t\}}] + \sum_{n=1}^{\infty} \int_0^t \mathbb{E}\left[\underbrace{Y(t)}_{=X_{n+1}-(t-y)} \mathbf{1}_{\{S_n\in(y,y+dy],N(t)=n\}} \right] \\
&= \mathbb{E}[(X_1 - t)\mathbf{1}_{\{X_1>t\}}] + \sum_{n=1}^{\infty} \int_0^t \mathbb{E}\left[(X_{n+1} - (t-y))\, \mathbf{1}_{\{S_n\in(y,y+dy]\}} \mathbf{1}_{\{X_{n+1}>t-y\}} \right] \\
&= \mathbb{E}[(X_1 - t)\mathbf{1}_{\{X_1>t\}}] + \int_0^t \mathbb{E}\left[X - (t-y)\mathbf{1}_{\{X>t-y\}} \right] \underbrace{\sum_{n=1}^{\infty} \mathbb{E}\left[\mathbf{1}_{\{S_n\in(y,y+dy]\}} \right]}_{=m(dy)} \\
&= \mathbb{E}[(X_1 - t)\mathbf{1}_{\{X_1>t\}}] + \int_0^t \mathbb{E}\left[X - (t-y)\mathbf{1}_{\{X>t-y\}} \right] m(dy)
\end{aligned}$$

The first term in the above sum converges to 0 as $t \to \infty$ since $\mathbb{E}[X_1]$ is finite. We can write $h(t-y) = X - (t-y)\mathbf{1}_{\{X>t-y\}}$ and use the Key

Renewal Theorem for the second term. If we do that we obtain the limit as:

$$\begin{aligned}
\mathbb{E}[Y(t)] &\xrightarrow{t\to\infty} = \frac{1}{\mu}\int_0^\infty h(s)ds = \frac{1}{\mu}\int_0^\infty \mathbb{E}[(X-s)\mathbf{1}_{\{X>s\}}]ds \\
&= \frac{1}{\mu}\int_0^\infty \left[\int_s^\infty (x-s)dF(x)\right] ds \\
\text{Fubini} &= \frac{1}{\mu}\int_0^\infty \int_0^x (x-s)dsdF(x) \\
&= \frac{1}{\mu}\int_0^\infty -\frac{(x-s)^2}{2}\Big|_0^x dF(x) \\
&= \frac{1}{\mu}\int_0^\infty \frac{x^2}{2}dF(x) = \frac{\mathbb{E}[X^2]}{2\mu}
\end{aligned}$$

□

Corollary 4.30. If $\mathbb{E}[X^2] < \infty$ and F non-lattice then (for the undelayed renewal process)

$$m(t) - \frac{t}{\mu} \xrightarrow{t\to\infty} \frac{\mathbb{E}[X^2]}{2\mu^2} - 1$$

PROOF. Note that we have shown that $\mathbb{E}[S_{N(t)+1}] = \mu \cdot (m(t)+1)$ However:

$$\mathbb{E}[t + Y(t)] = t + \mathbb{E}[Y(t)] \to t + \frac{\mathbb{E}[X^2]}{2\mu}$$

Since $S_{N(t)+1} = t + Y(t)$ we obtain:

$$m(t) + 1 \to \frac{t}{\mu} + \frac{\mathbb{E}[X^2]}{2\mu^2} \Rightarrow m(t) - \frac{t}{\mu} \to \frac{\mathbb{E}[X^2]}{2\mu^2} - 1$$

□

Example 4.31. Let $X_1, X_2 \ldots$ iid $U[0,1]$. Then $\mu = \frac{1}{2}$, $\mathbb{E}[X^2] = \frac{1}{3}$

Then the corollary says for $t = 100$

$$\begin{aligned}
m(100) &\sim \frac{100}{m} + \frac{\mathbb{E}[X^2]}{2\mu^2} - 1 \\
&= \frac{100}{\frac{1}{2}} + \frac{\frac{1}{3}}{2\cdot(\frac{1}{2})^2} - 1 \quad \leftarrow \text{better approximation} \\
&= 199\frac{1}{3} \quad \text{(probably very accurate)}
\end{aligned}$$

4.5.2. Renewal Reward Process. Consider iid pairs: $(X_1, R_1), (X_2, R_2) \dots$
Story: At time $S_n = \sum_{i=1}^{n} X_i$ you get a reward R_n. Assume that $X_i \geq 0$, $\mathbb{E}[X_i] = \mu < \infty$, $\mathbb{E}[R_i] < \infty$.
Let $R_t = \sum_{i=1}^{N(t)} R_i$ the total reward up to time t

Theorem 4.32. *Two results:*

(1)

$$\frac{R(t)}{t} \xrightarrow{a.s.} \frac{\mathbb{E}[R]}{\mu} \qquad as\ t \to \infty$$

(2)

$$\frac{\mathbb{E}[R(t)]}{t} \to \frac{\mathbb{E}[R]}{\mu} \qquad as\ t \to \infty$$

Proof. **Part (1):** We have:

$$\frac{R(t)}{t} = \frac{1}{t} \sum_{i=1}^{N(t)} R_i = \frac{\sum_{i=1}^{N(t)} R_i}{N(t)} \, \frac{N(t)}{t}$$

The first term in the product above converges to the $\mathbb{E}[R]$ using the strong law of large numbers and the second term converges to $1/\mu$ by the renewal SLLN. Therefore, we get the result in part (1).
Part (2): We have:
Using Wald for $N(t) + 1$ which is a stopping time,

$$\begin{aligned}\mathbb{E}[R(t)] = \mathbb{E}[\sum_{i=1}^{N(t)} R_i] &= \mathbb{E}[\sum_{i=1}^{N(t)+1} R_i] - \mathbb{E}[R_{N(t)+1}] \\ &= \mathbb{E}[N(t)+1]\mathbb{E}[R_n] - \mathbb{E}[R_{N(t)+1}] = (m(t)+1)\,\mathbb{E}(R) - \mathbb{E}[R_{N(t)+1}]\end{aligned}$$

This implies dividing with t and taking the limit as $t \to \infty$:

$$\frac{\mathbb{E}[R(t)]}{t} = \frac{(m(t)+1)}{t}\mathbb{E}(R) - \frac{\mathbb{E}[R_{N(t)+1}]}{t} \xrightarrow{t\to\infty} \frac{\mathbb{E}(R)}{\mu} - \lim_{t\to\infty} \frac{\mathbb{E}[R_{N(t)+1}]}{t},$$

where we used the elementary renewal theorem for the first term. To complete the proof we have to show that $\lim_{t\to\infty} \mathbb{E}[R_{N(t)+1}]/t = 0$. We have:

$$\begin{aligned}
\mathbb{E}[R_{N(t)+1}] &= \mathbb{E}[R_1\mathbf{1}_{\{X_1>t\}}] + \sum_{n=1}^{\infty}\int_0^t \mathbb{E}\left[R_{n+1}\mathbf{1}_{\{X_{n+1}>t-y, S_n\in(y,y+dy], N(t)=n\}}\right] \\
&= \mathbb{E}[R_1\mathbf{1}_{\{X_1>t\}}] + \int_0^t \sum_{n=1}^{\infty}\mathbb{E}\left[R_{n+1}\mathbf{1}_{\{X_{n+1}>t-y\}}\right]\mathbb{E}\left[\mathbf{1}_{\{S_n\in(y,y+dy]\}}\right] \\
&= \mathbb{E}[R_1\mathbf{1}_{\{X_1>t\}}] + \int_0^t \mathbb{E}\left[R_2\mathbf{1}_{\{X_2>t-y\}}\right]\sum_{n=1}^{\infty}\mathbb{E}\left[\mathbf{1}_{\{S_n\in(y,y+dy]\}}\right] \\
&= \mathbb{E}[R_1\mathbf{1}_{\{X_1>t\}}] + \int_0^t h(t-y)m(dy),
\end{aligned}$$

where we denoted $h(t-y) = \mathbb{E}\left[R_2\mathbf{1}_{\{X_2>t-y\}}\right]$, to apply the KRT. The first term converges to 0 as $t\to\infty$ (justify), and we obtain the limit:

$$\lim_{t\to\infty}\mathbb{E}[R_{N(t)+1}] = \frac{1}{\mu}\int_0^\infty h(t)dt = \frac{1}{\mu}\int_0^\infty \mathbb{E}\left[R_2\mathbf{1}_{\{X_2>t\}}\right] < \frac{\mathbb{E}(R)}{\mu} < \infty$$

Thus, dividing with t and taking the limit we obtain 0, which finishes the proof. □

4.6. The Renewal Equation. Convolutions.

Often the quantity of interest in renewal theory $Z(t)$ satisfies an equation of the form:

$$Z(t) = z(t) + \int_0^t Z(t-y)F(dy)$$

where $F(t)$ = c.d.f. of interarriaval time, and $z(t)$ is the some known function with the properties:

- $z(t) = 0$ if $t < 0$
- z bounded on finite interval

An equation of this type is called a **renewal equation**

Example 4.33. $m(t)$ satisfies:

$$m(t) = F(t) + \int_0^t m(t-y)F(dy)$$

Example 4.34. $P(Y(t) > x)$:

$$P(Y(t) > x) = \bar{F}(t+x) + \int_0^t P(Y(t-y) > x)F(dy)$$

Example 4.35. $\mathbb{E}[Y(t)]$:

$$\mathbb{E}[Y(t)] = \mathbb{E}[X_1 - t]I_{\{X_1 > t\}} + \int_0^t \mathbb{E}[Y(t-y)]dF(y)$$

The next theorem will provide a way to solve the renewal equation.

Theorem 4.36. *If $F(0-) = 0$, $F(0) < 1$, $z(t)$ is bounded on finite intervals and $z(t) = 0$ for $t < 0$ then the renewal equation*

$$Z(t) = z(t) + \int_0^t Z(t-s)dF(s)$$

has a unique solution, bounded on finite intervals given by

$$Z(t) = z(t) * m_0(t) = \int_0^t z(t-y)m_0(dy) = \sum_{n=0}^{\infty} \int_0^t z(t-y)dF_n(y)$$

where

$$m_0(t) = \sum_{n=0}^{\infty} F_n(t) = \sum_{n=0}^{\infty} P(S_n \leq t)$$
$$F_n(t) = \underbrace{F * F \ldots * F}_{n \text{ times}}, \qquad \text{with } S_0 = 0$$

Properties of Convolution Let F, G c.d.f.'s with $F(0-) = G(0-) = 0$, z as in the theorem. Then:

(1) $F * G = G * F$
(2) z*(F*G)=(z*F)*G
(3) z*(F+G)=z*F+z*G
(4) If G has density g then $F * G$ has density $g * F = \int_0^t g(t-y)F(dy)$

Proof of the theorem on renewal equation. **Part 1.** Existence of the solution.

$$\begin{aligned} z * m_0(t) &= \sum_{n=0}^{\infty} z * F_n(t) \\ &= z * F_0(t) + \sum_{n=1}^{\infty} z * F_n(t) \\ &= z(t) * F_0 + \left[\sum_{n=0}^{\infty} z * F_n(t)\right] * F(t) \\ &= z(t) + (z * m_0) * F(t) \\ &= z(t) + \int_0^t (z * m_0)(t-s) dF(s) \end{aligned}$$

Note that we used the fact $F_0(t) = P(S_0 \leq t) = \mathbf{1}_{\{t \geq 0\}}$

This shows that $z * m_0$ is a solution for the renewal equation.

Part 2. Uniqueness:

Assume that there exist $Z_1(t)$ and $Z_2(t)$ 2 solutions of the renewal equation. Let $V(t) = (Z_1 - Z_2)(t)$. By definition $V(t)$ should also solve the renewal equation, i.e.,

$$\begin{aligned} V(t) &= (Z_1 - Z_2)(t) \\ &= z(t) + \int_0^t Z_1(t-s) dF(s) - z(t) - \int_0^t Z_2(t-s) dF(s) \\ &= \int_0^t V(t-s) dF(s) = V * F(t) \end{aligned}$$

Repeat the argument:

$$V(t) = V * F(t) = V * F_2(t) = \ldots = V * F_k(t), \quad \forall k$$

which implies:

$$\begin{aligned} V(t) &= \int_0^t V(t-y) F_k(dy) \\ &\leq \sup_{0 \leq s \leq t} V(s) \int_0^t dF_k(s) \\ &= \sup_{0 \leq s \leq t} V(s) F_k(t) \xrightarrow{k \to \infty} 0 \end{aligned}$$

Because $F_k(t) = P(X_1 + X_2 + \ldots + X_k \leq t) \xrightarrow{k \to \infty} 0$, $\forall t$ fixed. (CLT or SLLN) $\square$

Theorem 4.37. *(true for both lattice and non-lattice case) If X_1 has distribution*

$$P(X_1 > x) = \int_0^\infty \frac{1}{\mu}\overline{F}(y)dy \stackrel{def}{=} F_e(x)$$

This is called the equilibrium distribution; the process with the delay X_1 having this distribution is called the equilibrium renewal process. Let

$$m_D(t) \stackrel{def}{=} \sum_{n=1}^{\infty} P(S_n \leq t) = \sum_{n=0}^{\infty} F_e * F_n(t),$$

and $Y_D(t)$ be the residual lifetime at t for the delayed process. Then:

(1) $m_D(t) = \frac{t}{\mu}$
(2) $P(Y_D(t) > x) = \overline{F}_e(x)$ *for all* $t > 0$
(3) $\{N_D(t)\}_t$ *has stationary increments.*

Proof. **Part 1.**

$$\begin{aligned} m_D(t) &= F_e(t) + \left(F_e * \sum_{n=1}^{\infty} F_n\right)(t) \\ &= F_e(t) + \left(F_e * \sum_{n=0}^{\infty} F_n\right) * F(t) \\ &= F_e(t) + m_D(t) * F(t) \end{aligned}$$

which implies that $m_D(t)$ solves a renewal equation with $z(t) = F_e(t)$

If we show that $\frac{t}{\mu}$ also solves the renewal equation with the same $z(t)$, we are done.

Check yourself that $h(t) = \frac{t}{\mu}\mathbf{1}_{\{t>0\}}$ also solves the same renewal equation. By uniqueness of the solution we are done.

Part 2 We have using the usual renewal argument:

$$
\begin{aligned}
P(Y_D(t) > x) &= \mathbf{P}(X_1 > t + x) + \int_0^t \overline{F}(t - y + x) m_D(dy) \\
(\text{From (i)} \Longrightarrow) &= \overline{F}_e(t + x) + \int_0^t \overline{F}(t - y + x) \frac{dy}{\mu} \\
&= \int_{t+x}^{\infty} \frac{1}{\mu} \overline{F}(y) dy + \int_0^t \frac{1}{\mu} \overline{F}(t - y + x) dy \\
(\text{c.v. } v = t - y + x) &= \int_{t+x}^{\infty} \frac{1}{\mu} \overline{F}(y) dy - \int_{t+x}^{x} \frac{1}{\mu} \overline{F}(v) dv \\
&= \int_x^{\infty} \frac{1}{\mu} \overline{F}(y) dy \\
&= \overline{F}_e(x) \quad \Longrightarrow \text{ DONE.}
\end{aligned}
$$

Part 3. This part follows from part (2) using the fact that $N_D(t + s) - N_D(s)$ is the number of renewals in a time interval length t of a delayed renewal process. □

CHAPTER 5

Special Chapter about some applications of the notions learned thus far.

At this point we are going to stop and look to some interesting examples of discrete processes.

5.1. Random Walk on integers in $\mathbb{R}^d$

Let $\vec{X}_k = (X_k^{(1)}, X_k^{(2)} \dots X_k^{(d)}) \in \mathbb{R}^d$ a random vector. Each $X_k^{(i)}$ is independent of the others and it is ± 1 each with probability $\frac{1}{2}$.

$$\text{Then } \vec{S}_n = \sum_{k=1}^{n} \vec{X}_k \quad \text{is a } d\text{-dimensional random walk}$$

Remark 5.1. The sum above is done componentwise, it is not the regular summation of the vectors. I use the notation $\vec{X}$ to symbolize the fact that X has more than one dimensions nothing more, there is no origin, directional angle or size involved in the notation.

We will talk next about some common questions regularly asked about this process.

Question: Once started from $(0, 0, \dots, 0)$, would the process come back to $\vec{0}$? OR is $S_n = (0, 0, \dots, 0)$ for some n?

Answer: For n odd, $P(S_n = \vec{0}) = 0$. For n even say equal to $2k$

$$\begin{aligned} P(\vec{S}_{2k} = \vec{0}) &= P\left(S_{2k}^{(1)} = 0, S_{2k}^{(2)} = 0, \dots S_{2k}^{(d)} = 0\right) \\ &= \left[P\left(S_{2k}^{(i)} = 0\right)\right]^d \end{aligned}$$

□

Claim: $\left[P\left(S_{2k}^{(i)} = 0\right)\right]^d = \left(\frac{\text{constant}}{\sqrt{k}}\right)^d$

Why?

Note that we have a total of $2k$ steps and $S_{2k}^{(i)}$ is now 1-dimensional. To get back to 0 once you start from it you need k steps up (values

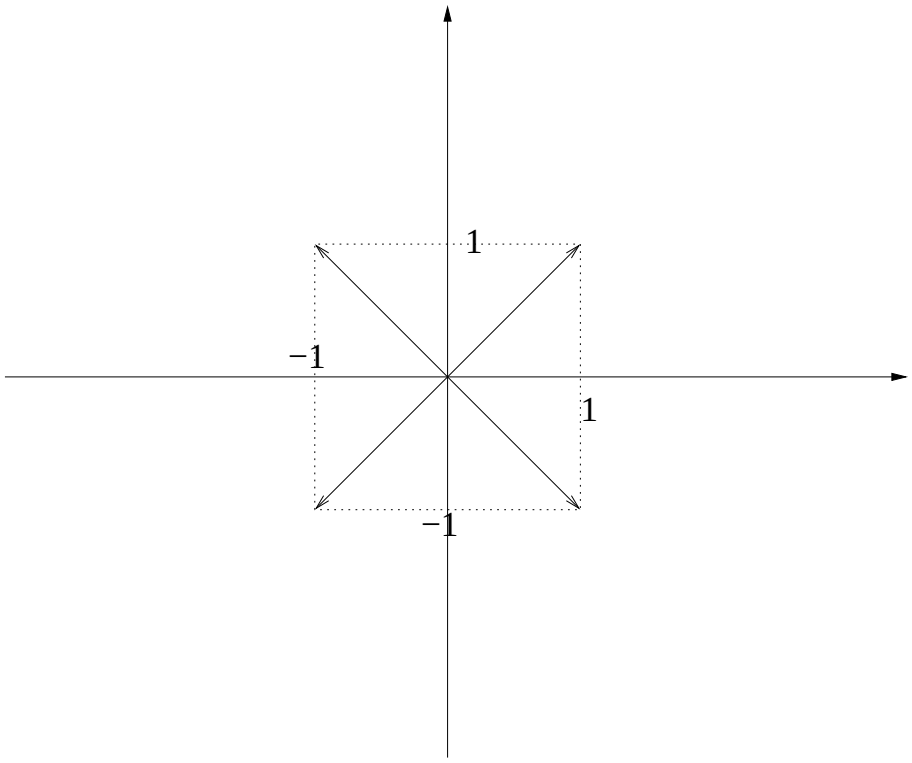

FIGURE 1. Random walk in 2-dimensions. Possible values for the first jump each with probability 1/4

of 1) and k steps down (values of -1). But then the number of such paths is:

$$2k \text{ total steps} \Rightarrow \begin{cases} k \text{ forward steps} \\ k \text{ back steps} \end{cases} \quad \rightarrow \binom{2k}{k}$$

Probability of any such path is $(\frac{1}{2})^{2k}$, which implies

$$P\left(S_{2k}^{(i)} = 0\right) = \binom{2k}{k}\left(\frac{1}{2}\right)^{2k}$$

Now use Stirling's formula for the combinatorial term, (i.e., $n! \sim \sqrt{2\pi n}\, n^n e^{-n}$):

$$\begin{aligned} P\left(S_{2k}^{(i)} = 0\right) &= \frac{(2k)!}{(k!)^2}\left(\frac{1}{2}\right)^{2k} \\ &= \frac{\sqrt{2\pi 2k}(2k)^{2k}e^{-2k}}{(\sqrt{2\pi k}k^k e^{-k})^2}\left(\frac{1}{2}\right)^{2k} \\ &= \frac{2\sqrt{\pi k}2^{2k}(k^{2k}e^{-2k})}{2\pi k(k^{2k}e^{-2k})}\left(\frac{1}{2}\right)^{2k} \\ &= \frac{1}{\sqrt{\pi k}} = \frac{\text{constant}}{\sqrt{k}} \end{aligned}$$

which proves the claim.

Theorem 5.2 (Polya)**.** *If $d = 1, 2$ then you come back to $\vec{0}$ infinitely often. If $d \geq 3$ eventually you never come back to $\vec{0}$.*

Proof. Let $p_d = \mathbf{P}\{\text{You never come back to } \vec{0} \textbf{ at all} \text{ in } \mathbb{R}^d\}$. Then:

$$\sum_{n=1}^{\infty} \mathbf{1}_{\{\vec{S}_n = \vec{0}\}} = \{\text{number of times you return to } \vec{0}\}$$

then $\sum_{n=1}^{\infty} \mathbf{1}_{\{\vec{S}_n=\vec{0}\}}$ is a Geometric(p_d) (number of failures) random variable. This is clear if you consider coming back to $\vec{0}$ a failure and not coming back a success. Therefore we can write:

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \mathbf{1}_{\{\vec{S}_n=\vec{0}\}}\right] = \frac{1}{p_d} - 1$$

Using Fubini and the previous claim:

$$\begin{aligned} \mathbb{E}\left[\sum_{n=1}^{\infty} \mathbf{1}_{\{\vec{S}_n=\vec{0}\}}\right] &= \sum_{n=1}^{\infty}\left(\mathbb{E}[\mathbf{1}_{\{\vec{S}_n=\vec{0}\}}]\right) = \sum_{n=1}^{\infty} P\left(\vec{S}_n = \vec{0}\right) = \sum_{\substack{n=2k \\ k=1}}^{\infty} \frac{c^d}{k^{\frac{d}{2}}} \\ &= c^d \sum_{k=1}^{\infty} \frac{1}{k^{\frac{d}{2}}} \end{aligned}$$

Therefore:

$$\frac{1-p_d}{p_d} = \begin{cases} < \infty & \text{if } d \leq 2 \\ \infty & \text{if } d > 2 \end{cases}$$

which implies

$$p_d > 0 \Leftrightarrow d \geq 3$$
$$p_d = 0 \Leftrightarrow d = 1, 2$$

As a conclusion when $d = 1, 2$ the number of visits to 0 is ∞. When $d >= 3$ the number of visits is finite a.s. which means that eventually you will drift to infinity. Moral: Do not get drunk while driving a spaceship. □

Remark 5.3. Considering a renewal event the event that the random walk returns to the origin $\vec{0}$ we can easily see that the random walk produces a renewal process.

5.2. Age dependent Branching processes

Story: Let F to be the lifetime distribution, with $F(0) = 0$,
P_j to be the probability that at death we get exactly j offsprings, $j = 0, 1, 2 \ldots$.
Each offspring then acts independently of others and produce their own offspring according to P_j, and so on and so forth.

Let $X(t)$ denote the number of organisms alive at time t. $\{X\}_{t>0}$ is called an age dependent branching process with $X(0) = 1$.

Quantity of interest : $M(t) = \mathbb{E}[X(t)]$

$$m = \sum_{j=0}^{\infty} jP_j = \text{number of offsprings (assumed to be } > 1)$$

Special case: If lifetimes are identically equal to 1, then

$$M(k) = m^k$$

In this particular situation $X(k)$ is called a Galton-Watson process (which is also a Markov Chain, as we will see later). This process was invented in 1873 by the people whose name it bears as result of a study initiated at the request of the crown to see if the aristocratic surnames were dying out in England of that time.

Remark : usually

$$\frac{X(k)}{m^k} \to Z$$

with Z a random variable finite a.s..

Theorem 5.4. *If $X(0) = 1$ and F is non-lattice then*

$$e^{-\alpha t} M(t) \xrightarrow{t\to\infty} \frac{m-1}{m^2 \alpha \int_0^\infty x e^{-\alpha x} dF(x)}$$

where $\alpha > 0$ is the solution of the equation: $\int_0^\infty e^{-\alpha t} dF(t) = \frac{1}{m}$

Remark 5.5. The theorem simply says that $M(t) \sim \text{constant} \cdot e^{-\alpha t}$

PROOF. Using $ds = (s, s + ds]$ and the renewal argument:

$$\begin{aligned} M(t) &= \mathbb{E}[X(t)] \\ &= \mathbb{E}[X(t)\mathbf{1}_{\{\text{1st life } > t\}}] + \mathbb{E}[X(t)\mathbf{1}_{\{\text{1st life} \le t\}}] \\ &= \mathbb{E}[1\mathbf{1}_{\{T_1 > t\}}] + \int_0^t \mathbb{E}[\underbrace{N}_{\text{Nr. of offsprings}} X(t-s)\mathbf{1}_{\{T_1 \in ds\}}] \\ &= P(T_1 > t) + \int_0^t m\mathbb{E}[X(t-s)\mathbf{1}_{\{T_1 \in ds\}}] \qquad \text{, note } m = \mathbb{E}[N] \\ &= \overline{F}(t) + \int_0^t mM(t-s)dF(s) \end{aligned}$$

This looks a lot like a renewal equation except for the m. To eliminate it multiply both sides by $e^{-\alpha t}$

$$\begin{aligned} \Rightarrow M(t)e^{-\alpha t} &= \overline{F}(t)e^{-\alpha t} + \int_0^t e^{-\alpha t} mM(t-s)dF(s) \\ &= \overline{F}(t)e^{-\alpha t} + \int_0^t e^{-\alpha(t-s)} M(t-s) \underbrace{me^{-\alpha s}dF(s)}_{=dG(s)} \end{aligned}$$

We denoted $dG(s) = me^{-\alpha s}dF(s)$, OR $G(t) = \int_0^t me^{-\alpha s}dF(s)$

G is a c.d.f. because α is a solution of $\int_0^\infty e^{-\alpha t}dF(t) = \frac{1}{m}$. Its definition implies that $G(0-) = 0$.

Thus, we obtain a renewal equation $Z(t) = z(t) + \int_0^t Z(t-s)dG(s)$, where:

$$\begin{aligned} Z(t) &= e^{-\alpha t}M(t) \\ z(t) &= e^{-\alpha t}\overline{F}(t) \end{aligned}$$

Recall that the unique solution is

$$\begin{aligned} Z(t) = z * m_0(t) &= \int_0^t z(t-s)m_0(ds) \qquad \text{(with } m_0 \text{ given by } \sum_{n=0}^{\infty} G_n(s)) \\ &\xrightarrow{\text{KRT}} \frac{1}{\mu_G}\int_0^\infty z(t)dt, \end{aligned}$$

provided that we can apply the Key Renewal Theorem. Looking back we see that $Z(t) \to \frac{1}{\mu_G}\int_0^\infty e^{-\alpha t}\overline{F}(t)dt$. Now let us calculate the limit, while at the same time showing that $z(t)$ is DRI. Using that $\overline{F}(t) = \int_t^\infty dF(x)$ we have:

$$\begin{aligned}
\int_0^\infty z(t)dt &= \int_0^\infty \left(\int_t^\infty dF(x)\right) e^{-\alpha t}dt \\
&= \int_0^\infty \left(\int_0^x e^{\alpha t}dt\right) dF(x) \qquad \text{(Fubini)} \\
&= \int_0^\infty \frac{1}{\alpha}(1-e^{-\alpha x})dF(x) \\
&= \frac{1}{\alpha}\left[\underbrace{\int_0^\infty dF(x)}_{=1} - \underbrace{\int_0^\infty e^{-\alpha x}dF(x)}_{=\frac{1}{m}\text{ (def. of }\alpha)}\right] \\
&= \frac{1}{\alpha}\left[1-\frac{1}{m}\right] < \infty
\end{aligned}$$

Thus, $z(t)$ is Riemann integrable, also it is decreasing and positive therefore it is DRI. All that remains is to calculate μ_G.

$$\mu_G = \int_0^\infty xdG(x) = \int_0^\infty xme^{-\alpha x}dF(x) = m\int_0^\infty xe^{-\alpha x}dF(x)$$

Hence,

$$Z(t) = e^{-\alpha t}M(t) \to \frac{\frac{1}{\alpha}\left(1-\frac{1}{m}\right)}{m\int_0^\infty \alpha e^{-\alpha x}dF(x)}$$

And a little algebra shows that this is exactly the formula we need to prove. □

Remark 5.6. What if $m < 1$? If $\exists\alpha < 0$ with $\int_0^\infty e^{-\alpha x}dF(x) = \frac{1}{m}$ and $e^{-\alpha x}\overline{F}(x)$ is DRI, then the same result is true.

In either case $m > 1$, $m < 1$; $\mu_G = \infty$ is possible and it will not change the answers.

Exercise 22. Question For a Branching Process what is the probability that $X(t) = 0$ eventually? (population dies out).
Think about this.
Guesses.

For $m < 1$, it is kind of obvious that $P(\text{Population dies out}) = 1$.

If $m = 1$, then $P(\text{Population dies out}) = 1$ except when the number of offsprings is exactly 1.

What if $m > 1$? $P(\text{Population dies out}) > 0$ iff $P(0 \text{ offsprings}) > 0$

CHAPTER 6

Markov Chains

In this chapter we start the study of the first and one of the most popular models for real life situations coming from probability.

6.1. Basic concepts of Markov Chains

6.1.1. Definition. Consider a set of outcomes $\mathcal{S}$ which is finite or countable. It is convenient to represent the set $\mathcal{S}$ as the nonnegative integers $\{0, 1, \ldots\}$. Consider a process $X = (X_1, X_2, \ldots, X_n, \ldots)$ whose components X_n take values in this set $\mathcal{S}$. We will say that the process X is in state $i \in \mathcal{S}$ at time n if $X_n = i$.

Consider a matrix

$$P = \begin{pmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \ldots \\ P_{1,0} & P_{1,1} & P_{1,2} & \ldots \\ \vdots & \vdots & \vdots & \\ P_{i,0} & P_{i,1} & P_{i,2} & \ldots \\ \vdots & \vdots & \vdots & \end{pmatrix},$$

with

$$\begin{cases} P_{i,j} \geq 0, \forall i, j \\ \sum_{j=0}^{\infty} P_{i,j} = 1, \forall i \text{ (i.e., rows sum to 1).} \end{cases}$$

Furthermore, let $\pi_0 = \{\pi_0(i)\}_{i=0}^{\infty}$ be a probability mass function on the elements of $\mathcal{S}$ (i.e., $\pi_0(i) \geq 0$ and $\sum_i \pi_0(i) = 1$).

Definition 6.1. A discrete time Markov chain on the state space $\mathcal{S}$, with initial distribution π_0 and and transition matrix P is a stochastic process $X = \{X_n\}_{n=0}^{\infty}$ with

$$\mathbf{P}\left(X_0 = i_0, \ldots, X_n = i_n\right) = \pi_0(i_0) \prod_{k=0}^{n-1} P_{i_k, i_{k+1}}$$

Note that with this definition p_{ij} can be interpreted as the transition probability of jumping from state i to state j in one step.

Theorem 6.2 (Markov Property)**.** *If $X = \{X_n\}_{n=0}^{\infty}$ is a discrete time Markov chain defined as above then:*

$$\mathbf{P}\left(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n\right) = \mathbf{P}\left(X_{n+1} = i_{n+1} | X_n = i_n\right) = P_{i_n, i_{n+1}}$$

PROOF. Exercise. Verify the theorem using the Definition 6.1. □

6.1.2. Examples of Markov chains. This section presents several simple examples.

Example 6.3. This is a very simple example. Let $X_1, X_2, \ldots, X_n, \ldots$ be iid random variables on $\mathbb{Z}$. Take the transition probabilities $p_{i,j} = \mathbf{P}(X_k = j)$, for any k. Then this forms a (rather trivial) Markov Chain on $\mathbb{Z}$ with transition probabilities not dependent on the current state (i.e. all rows in the transition matrix are identical).

Example 6.4 (Generalized random walk on $\mathbb{Z}$)**.** Let $\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_n, \ldots$ be again iid random variables on $\mathbb{Z}$ with some probability mass function: $p_m = \mathbf{P}(\Delta_k = m)$. Let the process X with components given by:

$$X_n = \sum_{k=0}^{n} \Delta_k$$

Note that unlike the simple random walk the generalized one can have jumps of any size. It is easy to see that X is a Markov chain on $\mathbb{Z}$ with initial distribution $\pi_0(m) = p_m$ and with transition probability matrix having elements:

$$P_{i,j} = \mathbf{P}(\Delta_k = j - i) = p_{j-i}.$$

As a special case if all the jumps are positive $\Delta_k \geq 0$ we obtain a renewal process.

Example 6.5 (Ehrenfest chain)**.** Suppose we have two boxes, box 1 with x balls and box 2 with $d - x$ balls (a total of d balls in both). At each step (time) we pick one of the balls at random and we transfer that ball from one box to another (see figure 1).

We consider the process $X = \{X_n\}_{n=0}^{\infty}$ with X_n given by the number of balls in box 1 at time n. We obviously have $X_n \in \mathcal{S} = \{0, 1, .., d\}$. Then X is a Markov chain on $\mathcal{S}$ with some initial distribution. The transition probabilities for the chain are:

$$\mathbf{P}(X_n = y | X_{n-1} = x) = P_{x,y} = \begin{cases} \frac{x}{d}, & \text{if } y = x - 1, \\ \frac{d-x}{d}, & \text{if } y = x + 1, \\ 0, & \text{else .} \end{cases}$$

Ehrenfest used this model to study the exchange of air molecules in two chambers connected by a small hole.

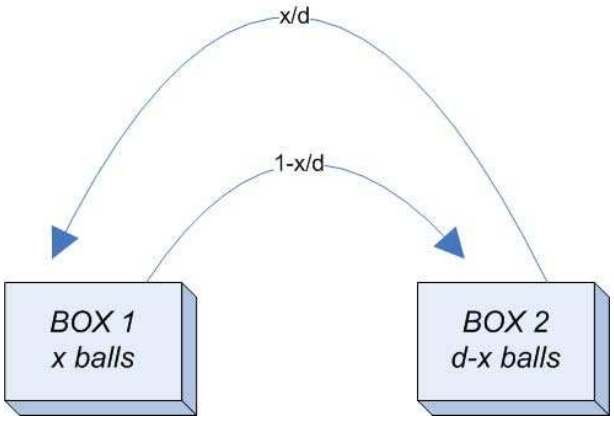

FIGURE 1. A graphical description of the Ehrenfest Chain

We also note that this is an example of a periodic chain (the proper definition to come) since the number of balls in the box 1 alternates between odd and even.

Example 6.6 (Gambler's ruin problem)**.** Suppose we deal with the case of a compulsive gambler that starts with some initial wealth K. At every step the gambler bets one dollar on some game that has probability p of winning and if he/she wins his wealth goes up by one dollar. Furthermore, assume that the gambler stops betting only when his/her wealth reaches either 0 or some upper limit fixed apriory $d > K$ (which may be ∞). Let $X_0 = K$ and let $X_k =$ Gambler's wealth after k such games. Then X_n is a Markov chain on $\mathcal{S} = \{1, 2, \ldots, d\}$, with initial distribution $\pi_0(i) = \mathbf{1}_{\{K\}}(i)$ and transition probabilities:

$$\text{If } 0 < i < d \text{ then } \begin{cases} P_{i,i+1} = p \\ P_{i,i-1} = 1 - p \\ P_{i,j} = 0, \qquad \text{for any other } j \end{cases}$$
$$P_{0,j} = \mathbf{1}_{\{0\}}(j); P_{d,j} = \mathbf{1}_{\{d\}}(j);$$

Note that this process can be interpreted as a simple (± 1) random walk with 0 and d absorbtion states (see diagram in figure **??**).

An interesting question (as with all random walks) is what is the probability of eventually hitting 0 (probability of eventual ruin).

Example 6.7 (Birth and Death chain)**.** This is a generalization of both previous examples. We have absorbtion states as in the example 6.6 and the probabilities of transition are state dependent as in example 6.5. Furthermore, the process can remain in the same state.

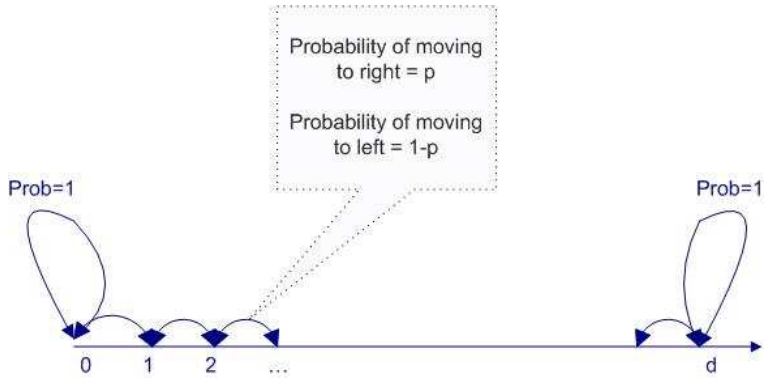

FIGURE 2. A graphical description of the Gambler's ruin problem

The story is that we start with a certain population type. We view the time between events as non-important so we take that the time step is 1 (thus creating a Markov chain). At any time step one event take place: either a birth or a death or nothing occurs. If X is the process denoting the size of the population then it is a Markov chain on $\mathcal{S} = \{0, 1, 2, \ldots\}$ with some initial distribution. We can see a scheme depicting it in figure 3, assuming that the population dies out if reaching size 0 (no immigration from outside). We could also put an upper cap on the population by creating another (semi) absorbing upper state.

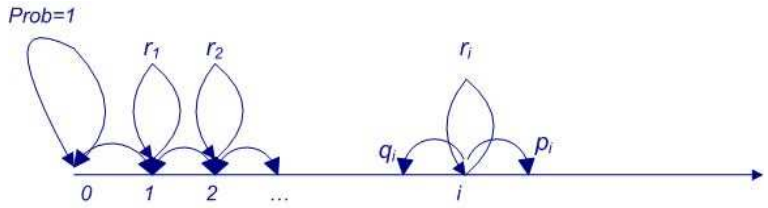

FIGURE 3. A graphical description of the Birth and Death Chain

We can write down the transition probabilities of this chain as:

$$\text{If } 0 < i \text{ then } \begin{cases} P_{i,i+1} = p_i \\ P_{i,i-1} = q_i \\ P_{i,i} = r_i \\ P_{i,j} = 0, \qquad \text{for any other } j \end{cases}$$

$$P_{0,j} = \mathbf{1}_{\{0\}}(j),$$

assuming that for any i we have $p_i + q_i + r_i = 1$.

Example 6.8 (A simple queuing process (G/G/1))**.** Assume that we have customers arriving at a service facility. They are processed by one service facility according to some general distribution. In queuing theory G/G/1 is a notation where the first letter denotes the arrival distribution (G stands for general), the second letter stands for service distribution (once again G means general) and the last number symbolizes the number of servers in the queuing system (in this case a single server). Let Δ_n denote the number of customers arriving during the n'th service. Let $p_m = \mathbf{P}(\Delta_n = m)$, $m = 0, 1, 2, \ldots$. In general this distribution will depend on the length of the n'th service and we will obtain a so called non-homogeneous Markov chain but for now let us make the (unrealistic) assumption that the distribution is the same for all n (symbolized above by the fact that p_m does not have an index n).

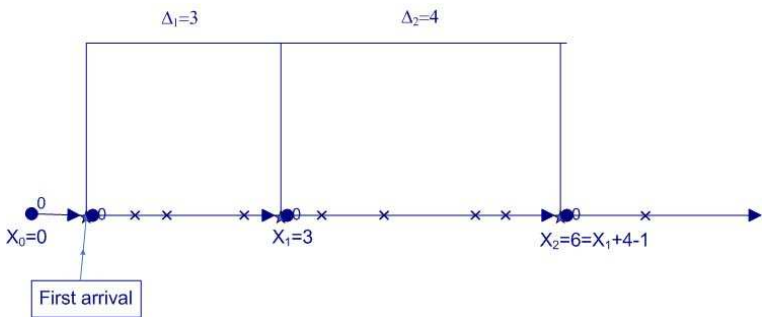

FIGURE 4. A realization of the simple queue system

Let X_{n+1} be the number of customers in queue after the n'th service. A typical realization is depicted in figure 4. Then we can write in general (think why):

$$X_{n+1} = X_n + \Delta_n - \mathbf{1}_{\{X_n>0\}}.$$

Therefore, $\{X_n\}$ is a Markov chain with state space $\mathbb{N} = \{0, 1, 2, \ldots\}$; with initial distribution $\pi_0(i) = \mathbf{1}_{\{0\}}(i)$; and with transition probabilities given by:

$$P_{i,j} = \mathbf{P}(\Delta_n = j - i + 1) = p_{j-i+1}, \text{ if } i \geq 1$$
$$P_{0,j} = p_j$$

Example 6.9 (Birth and catastrophe chain)**.** This is similar with the Birth and Death chain in Example 6.7 with the difference that there are no natural deaths, instead at any moment there is a probability that the entire population is wiped out. Furthermore, to not get stuck we assume that once at zero there is a certain probability for spontaneous

life. We can easily show that X the process of population size after step n is a Markov chain and its transition probabilities are given by:

$$P_{i,j} = \begin{cases} p_i, & \text{if } j = i+1 \\ q_i, & \text{if } j = 0 \\ 0, & \text{else} \end{cases} \quad ,\forall i \geq 0.$$

A simple example of such a process that we have already seen is the current age process $A(t)$ for a renewal process restricted to time in $\mathbb{Z}$.

Example 6.10 (Number of Mutant Genes in a population)**.** Simplified model.
Story: Suppose the existence of i mutant genes (from a total of d genes) in the n-th generation of a certain population. To get to the next generation:

(1) All genes duplicate, therefore there will be $2i$ mutant and $2d-2i$ normal genes
(2) d genes are randomly selected from the above possibilities.

If X_n denotes the number of mutant genes in the population after the n-th generation then $X = \{X_n\}_{n=1}^{\infty}$ is a Markov chain with state space $\mathcal{S} = \{1, 2, \ldots, d\}$ and with transition probabilities given by:

$$P_{i,j} = \frac{\binom{2i}{j}\binom{2d-2i}{d-j}}{\binom{2d}{d}} \quad ,\forall i, j \in \{1, 2, \ldots, d\}.$$

6.1.3. The Chapman-Kolmogorov equation.

Theorem 6.11. *Let $X = \{X_n\}_{n=1}^{\infty}$ be a Markov chain on $\mathcal{S} = \{1, 2, \ldots\}$, with some initial distribution and one step transition probability matrix P. Denote the n-step transition probability of going from state i to state j by $P_{i,j}^n$, i.e. $P_{i,j}^n = \mathbf{P}(X_{m+n} = j | X_m = i)$. Then we have the Chapman-Kolmogorov relation:*

$$P_{i,j}^{n+m} = \sum_{k \in \mathcal{S}} P_{i,k}^n P_{k,j}^m. \tag{6.1}$$

PROOF. The proof is simple:

$$\begin{aligned} P_{i,j}^{n+m} &= \mathbf{P}\left(X_{m+n} = j | X_0 = i\right) = \sum_{k\in\mathcal{S}} \mathbf{P}\left(X_{m+n} = j, X_n = k | X_0 = i\right) \\ &= \sum_{k\in\mathcal{S}} \mathbf{P}\left(X_{m+n} = j | X_n = k, X_0 = i\right) \mathbf{P}\left(X_n = k | X_0 = i\right) \\ &= \sum_{k\in\mathcal{S}} P_{k,j}^m P_{i,k}^n, \end{aligned}$$

where the first equality is the definition, for the second we used the law of total probability (exhaustive events: $\cup_{k\in\mathcal{S}}\{X_n = k\}$), third equality is just the multiplicative rule and finally the last follows from the Markov property (Theorem 6.2). □

Remark 6.12. This theorem also tells us that if $P^{(n)}$ is the n-step transition matrix then $P^{(n)} = P^n$ the n-th power of the 1-step transition matrix.

6.2. Exercises

1. Give the proof of the Markov property (Theorem 6.2).
2. Show that the Remark 6.12 is true.

Bibliography

[Bil95] Patrick Billingsley, *Probability and measure*, 3 ed., Wiley, 1995.

[GS01] Geoffrey Grimmett and David Stirzaker, *Probability and random processes*, 3 ed., Oxford University Press, 2001.

[KT75] Samuel Karlin and Howard M. Taylor, *A first course in stochastic processes*, 2 ed., Academic Press, 1975.