# PROBABILITY BASICS

JOHN E. HUTCHINSON

## Contents



## 1. Set Theoretic material

For a sequence $(A_n)$ of subsets of a set $X$ define

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geq 1}\bigcap_{m\geq n} A_m, \quad \limsup_{n\to\infty} A_n = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m. \tag{1}$$

Note

$$\bigcap_{m\geq n} A_m \uparrow \liminf_{n\to\infty} A_n, \quad \bigcup_{m\geq n} A_m \downarrow \limsup_{n\to\infty} A_n, \quad \text{as } n\to\infty. \tag{2}$$

In words,

$$x\in \liminf_{n\to\infty} A_n \iff x\in A_n \text{ eventually}, \quad x\in \limsup_{n\to\infty} A_n \iff x\in A_n \text{ i.o.} \tag{3}$$

In particular, $\liminf A_n$ is the smallest set detectible by tail events and $\limsup A_n$ is the largest.

Clearly,

$$(\liminf A_n)^c = \limsup A_n^c, \quad (\limsup A_n)^c = \liminf A_n^c. \tag{4}$$

That is

$$\neg(A_n \text{ eventually}) \iff A_n^c \text{ i.o.}, \quad \neg(A_n \text{ i.o.}) \iff A_n^c \text{ eventually}. \tag{5}$$

*Date*: October 15, 2010.

In terms of *indicator* or *characteristic* functions,

$$\mathbf{1}_{\liminf_n A_n} = \liminf_n \mathbf{1}_{A_n}, \quad \mathbf{1}_{\limsup_n A_n} = \limsup_n \mathbf{1}_{A_n}, \tag{6}$$

where on the right of each equality we are taking the liminf and limsup of a sequence of $\{0, 1\}$-valued *functions*. It follows from (3) and (6), since $\mathbf{1}_A(x)$ is either 0 or 1, that

$$x \in A_n \text{ eventually} \iff \prod_{m\geq n} \mathbf{1}_{A_m}(x) = 1 \text{ eventually} \iff \lim_n \prod_{m\geq n} \mathbf{1}_{A_m}(x) = 1, \tag{7}$$

$$x \in A_n \text{ i.o.} \iff \sum_n \mathbf{1}_{A_n}(x) = \infty. \tag{8}$$

If $X = \Omega$ is a sample space, the $A_n$ are interpreted as events and we write

$$A_n \text{ eventually} \iff \prod_{m\geq n} \mathbf{1}_{A_m} = 1 \text{ eventually} \iff \lim_n \prod_{m\geq n} \mathbf{1}_{A_m} = 1, \tag{9}$$

$$A_n \text{ i.o.} \iff \sum_n \mathbf{1}_{A_n} = \infty. \tag{10}$$

## 2. Weak Law of Large Numbers

The word "weak" refers to convergence in probability, as opposed to the later "strong" laws which give convergence a.e.

**Theorem 2.1.** *Suppose $(X_n)_{n\geq 1}$ are iid real R.V.'s with mean $\mu$ and finite variance. Then for any $\epsilon > 0$,*

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0.$$

*Proof.* By Chebyshev,

$$P\left(\left|\sum_{i=1}^{n}(X_i - \mu)\right| \geq n\epsilon\right) \leq \frac{\mathbb{E}\left(\sum_{i=1}^{n}(X_i - \mu)\right)^2}{n^2\epsilon^2} \leq \frac{n\,\mathbb{E}(X_1 - \mu)^2}{n^2\epsilon^2} \to 0, \quad \text{as } n \to \infty.$$

The result follows. □

## 3. Zero One Law

This follows trivially for i.o. events from the Borel-Cantelli Lemma, but is easy to show directly as follows. (See [Lam66, pp 37–41].)

Let $X_n$ be R.V.'s. The $\sigma$-algebras $\mathcal{B}(X_n, X_{n+1}, \dots)$ are decreasing as $n \to \infty$. The intersection $\mathcal{B}_\infty$ is called the *tail field* of the sequence.

It follows that $\mathcal{B}_n := \mathcal{B}(X_n)$ is independent of $\mathcal{B}_\infty$ for every $n$.

If the $X_n$ are independent then the following says the tail field is trivial probabilistically.

**Proposition 3.1** (Zero-One law)**.** *For independent events, any tail event has probability* 0 *or* 1.

*Proof.* $\mathcal{B}_\infty$ is independent of $\mathcal{B}_n$ for any $n$, and so is independent of itself. Hence $P(E)^2 = P(E)$ for any tail event $E$, which gives the result. □

## 4. Borel-Cantelli Lemma

*(Partly from some now removed web notes* [Che09]*)*

The main point in the following proposition for proving (1) is that $\mathbb{E}f < \infty$ implies $f < \infty$ a.s., where $f = \sum_n \mathbb{I}_{A_n}$.

The main point for proving (2) is that $\sum_n P(A_n) = \infty$ implies $P\left(\bigcap_{k\geq n} A_k^c\right) = 0$ (by algebra and independence), i.e. $P\left(\bigcup_{k\geq n} A_k\right) = 1$.

**Proposition 4.1.** *For events $A_1, A_2, \dots$*

(1) $\sum_n P(A_n) < \infty \implies P(A_n \text{ i.o.}) = 0$ ($\iff \neg A_n$ *eventually, a.s.*),
(2) $A_n$ *independent*, $\sum_n P(A_n) = \infty \implies P(A_n \text{ i.o.}) = 1$.

*In particular, if* $A_n$ *are independent then* $P(A_n \text{ i.o.}) = 1$ *or* $0$, ($\iff A_n$ *i.o., a.s.*)

*Proof.* For (1) assume $\sum_n P(A_n) < \infty$. Hence

$$\infty > \sum_n P(A_n) = \sum_n \mathbb{E}\mathbf{1}_{A_n} = \mathbb{E}\sum_n \mathbf{1}_{A_n},$$

The first equality is by the definitions and the second by the monotone convergence theorem. This implies (but is not equivalent to) $\sum_n \mathbf{1}_{A_n} < \infty$ a.s.

But

$$\sum_n \mathbf{1}_{A_n} < \infty \text{ a.s.} \iff P\Big(\sum_n \mathbf{1}_{A_n} = \infty\Big) = 0 \iff P(A_n \text{ i.o.}) = 0.$$

The first equivalence is by the definitions and the second by (8).

For (2) first note that if $t_k \in [0,1]$ then $1 - t_k \leq e^{-t_k}$. Hence $\sum_n P(A_n) = \infty$ implies $\prod_{k\geq n}(1 - P(A_k)) = 0$ for all $n$. Hence

$$0 = \lim_n \prod_{k\geq n} P(A_k^c) = \lim_n \prod_{k\geq n} \mathbb{E}\mathbf{1}_{A_k^c} = \lim_n \mathbb{E}\prod_{k\geq n} \mathbf{1}_{A_k^c} = \mathbb{E}\lim_n \prod_{k\geq n} \mathbf{1}_{A_k^c} = \mathbb{E}\liminf_n \mathbf{1}_{A_n^c}.$$

The second equality is by definitions, the third by independence and D.C.T., the fourth by D.C.T., the last since $\mathbf{1}_{A_n^c}$ is zero or 1.

Hence $P(\liminf A_n^c) = 0$, hence $P(\limsup A_n) = 1$ by (4), i.e. $P(A_n \text{ i.o.}) = 1$ by (3).

*Alternatively*, rewriting the proof of (2) a little we have:

$\sum_n P(A_n) = \infty$ implies (as above) for all $n$ that $\prod_{k\geq n}(1-P(A_k)) = 0$, i.e. $\prod_{k\geq n} P(A_k^c) = 0$. By independence (see below to be more rigorous), one has for all $n$ that $P(\bigcap_{k\geq n} A_k^c) = 0$, i.e. $P(\bigcup_{k\geq n} A_k) = 1$. Hence $P(\limsup A_n) = 1$, i.e. $P(A_n \text{ i.o.}) = 1$.

(More precisely for the independence step above, one has

$$P\left(\bigcap_{k\geq n} A_k^c\right) = \lim_{m\to\infty} P\left(\bigcap_{k=n}^m A_k^c\right) = \lim_{m\to\infty}\prod_{k=n}^m P\left(A_k^c\right) = 0,$$

where the first equality is by having a decreasing sequence of sets and the second by independence.) □

## 5. Applications of Borel-Cantelli Lemma

### 5.1. Convergence to zero a.s.

**Proposition 5.1.** *Suppose* $(X_n)_{n\geq 1}$ *are real r.v.'s and* $\sum_n P(|X_n| > \epsilon) < \infty$ *if* $\epsilon > 0$. *Then* $X_n(\omega) \to 0$ *a.s.*

*Proof.* By Borel-Cantelli with $\epsilon = 1/k$ for any positive integer $k$, $|X_n| \leq 1/k$ eventually a.s. Hence $X_n \to 0$ a.s. □

5.2. **Geometric Random Variables.** Consider a random sequence $(X_n)_{n\geq 1}$ where $X_n \in \{0,1\}$.

Let $n_k$ be the $k$th occurrence of 1 for $k \geq 1$ and let $n_0 = 0$.

Let $\ell_k = n_k - n_{k-1} - 1$ for $k \geq 1$. That is, $\ell_k$ is the length of the $k$th run of 0's.

We want an a.s. eventual upper bound on the growth of $\ell_k$. Optimally, we want a sequence $\phi_k$ and a constant $\theta_0$ such that

$$\limsup_{k\to\infty} \frac{\ell_k}{\phi_k} = \theta_0. \tag{11}$$

**Step A**: Suppose one can show

$$\sum_{k\geq 1} P(\ell_k > \theta\,\phi_k) < \infty.$$

Then by Borel-Cantelli,

$$\ell_k \le \theta\,\phi_k \text{ eventually a.s.}, \quad \Longrightarrow \limsup_k \frac{\ell_k}{\phi_k} \le \theta \text{ a.s.}$$

If this is true for every $\theta > \theta_0$ then

$$\limsup_k \frac{\ell_k}{\phi_k} \le \theta_0 \text{ a.s.}$$

**Step C**: Suppose one can show

$$\sum_{k\ge 1} P(\ell_k \ge \theta_0\phi_k) = \infty,$$

and the $\ell_k$ are independent. Then by Borel-Cantelli,

$$\ell_k \ge \theta_0\phi_k \text{ i.o., a.s.}, \quad \Longrightarrow \limsup_k \frac{\ell_k}{\phi_k} \ge \theta_0 \text{ a.s.}$$

**Step B**: Now suppose $X_n$ are iid with $P(X_n = 0) = p$. Note that the $\ell_k$ are iid and that $P(\ell_k \ge m) = p^m$ if $m$ is a positive integer.

In order to obtain a lim sup estimate as in (11) we need to consider a suitable sequence $\phi_k$ and investigate for which $\theta$

$$\sum_k P\left(\frac{\ell_k}{\phi_k} \ge \theta\right) \sim \sum_k p^{\theta\phi_k}$$

converges, and for which $\theta$ it diverges.

Motivated by

$$\sum_k \frac{1}{k^{1+\epsilon}} \quad \begin{cases} = \infty & \text{if } \epsilon = 0 \\ < \infty & \text{if } \epsilon > 0 \end{cases},$$

we consider the one parameter family of sequences $(\phi_k)_{k\ge 1}$ defined by

$$p^{\phi_k} = \frac{1}{k^{1+\epsilon}}, \quad \text{i.e. } \phi_k = \frac{(1+\epsilon)\log k}{\log(1/p)}$$

for $\epsilon \ge 0$. (The previous factor $\theta$ corresponds to the term $1+\epsilon$.)

From Steps A and B,

$$\limsup_k \frac{\ell_k}{\phi_k} = 1 \quad \text{if } \phi_k = \frac{\log k}{\log(1/p)}.$$

That is

$$\limsup_k \frac{\ell_k}{\log k} = \log\left(\frac{1}{p}\right). \tag{12}$$

## 6. Strong Law of Large Numbers (moment restriction)

**Theorem 6.1.** *Suppose $(X_n)_{n\ge 1}$ are iid real R.V.'s with mean $\mu$ and finite fourth moment. Then*

$$\frac{X_1 + \cdots + X_n}{n} \to \mu \quad a.s.$$

*Proof.* Let $\mathbb{E}(X_1 - \mu)^2 = \sigma^2$. By Chebyshev,

$$P\left(\left|\sum_{i=1}^n (X_i - \mu)\right| \ge n\epsilon\right) \le \frac{\mathbb{E}\left(\sum_{i=1}^n (X_i-\mu)\right)^4}{n^4\epsilon^4} \le \frac{n\,\mathbb{E}(X_1-\mu)^4 + 6\binom{n}{2}\sigma^4}{n^4\epsilon^4} \le \frac{Cn^2}{n^4\epsilon^4}.$$

Using the Borel-Cantelli lemma as in Proposition 5.1, the result follows. □

## 7. Strong Law of Large Numbers (Kolmogorov)

In this section we drop the fourth moment requirement.

7.1. **Maximal Inequality; a.s. Convergence.** We first prove the following "maximal" inequality due to Kolmogorov. For this result it is helpful to think of $S_i = X_1 + \cdots + X_i$ as being the $i$th position in a type of random walk beginning from the origin. The theorem gives an upper bound on the probability that the walk escapes $(-a, a)$ within the first $n$ steps. The same result and proof applies in $\mathbb{R}^d$ with $B_a(0)$ instead of $(-a, a)$.

Note that the result reduces to a simple version of Chebyshev's inequality in case $n = 1$.

**Theorem 7.1.** *Let $X_1, \ldots, X_n$ be independent with zero means and variances $\sigma_n^2$. Then for any $a > 0$,*

$$P\left(\max_{1\le i\le n} |X_1 + \cdots + X_i| \ge a\right) \le \frac{\sum_{i=1}^n \sigma_i^2}{a^2}.$$

*Proof.* Let $S_i = X_1 + \cdots + X_i$.

Let

$$A = \{\, S_i \notin (-a, a) \text{ for some } i \in \{1, \ldots, n\} \,\}$$
$$A_j = \{\, S_1, \ldots, S_{j-1} \in (-a, a),\ S_j \notin (-a, a) \,\}.$$

Note $A = \bigcup_j A_j$ and this is a disjoint union.

It follows

$$\begin{aligned}
\sum_{i=1}^n \sigma_i^2 = \mathbb{E}(S_n^2) \ge \mathbb{E}(S_n^2\,\mathbb{I}_A) &= \sum_j \mathbb{E}(S_n^2\,\mathbb{I}_{A_j}) \\
&= \sum_j \mathbb{E}\left((S_j + S_n - S_j)^2\,\mathbb{I}_{A_j}\right) \\
&= \sum_j \left(\mathbb{E}(S_j^2\,\mathbb{I}_{A_j}) + \mathbb{E}((S_n - S_j)^2\,\mathbb{I}_{A_j})\right) \quad (S_j\,\mathbb{I}_{A_j} \text{ and } S_n - S_j \text{ are independent}) \\
&\ge a^2 P(A_j) = a^2 P(A).
\end{aligned}$$

This is the required result. □

**Theorem 7.2.** *Let $(X_n)_{n\ge1}$ be independent R.V.'s with zero means and variances $\sigma_n^2$. Suppose $\sum_n \sigma_n^2 < \infty$. Then $\sum_n X_n$ converges a.s.*

*Proof.* Suppose $\epsilon > 0$. From the previous theorem,

$$P\left(\max_{m<i\le n} |X_m + \cdots + X_i| \ge \epsilon\right) \le \frac{\sum_{i=m}^n \sigma_i^2}{\epsilon^2} \le \frac{\sum_{i\ge m} \sigma_i^2}{\epsilon^2}.$$

Since this is true for any $n$,

$$P\left(\sup_{i>m} |X_m + \cdots + X_i| \ge \epsilon\right) \le \frac{\sum_{i\ge m} \sigma_i^2}{\epsilon^2} \to 0 \text{ as } m \to \infty.$$

Hence, almost surely, the sequence of partial sums eventually oscillates by at most $\epsilon$. Taking a sequence $\epsilon_k \to 0$, the sequence of partial sums converges a.s. □

7.2. **Kronecker's Lemma; SLLN for differing distributions.** We first need

**Theorem 7.3** (Kronecker's Lemma)**.**

$$\sum_j \frac{a_j}{j} \textit{ converges} \quad \Longrightarrow \quad \frac{a_1 + \cdots + a_n}{n} \to 0.$$

*Proof.* Let

$$s_n = \sum_{j=1}^n \frac{a_j}{j}.$$

Then

$$\begin{aligned}\sum_{j=1}^{n} a_j &= \sum_{j=1}^{n} j s_j \\ &= s_1 + 2(s_2 - s_1) + 3(s_3 - s_2) + \cdots + n(s_n - s_{n-1}) \\ &= -(s_1 + s_2 + \cdots + s_{n-1}) + n s_n\end{aligned}$$

Then

$$\frac{a_1 + \cdots + a_n}{n} = s_n - \frac{s_1 + s_2 + \cdots + s_{n-1}}{n}.$$

We know $s_n \to x$, say. It follows easily that $(s_1 + s_2 + \cdots + s_{n-1})/n \to x$. This gives the result. □

A more general version of the above is in [Ash00, p236].

Of course the following applies to any finite mean $\mu$ by considering $X_n - \mu$.

**Theorem 7.4** (Kolmogorov)**.** *Let $(X_n)_{n\geq 1}$ be independent R.V.'s with zero means and variances $\sigma_n^2$. Suppose $\sum_n \sigma_n^2/n^2 < \infty$. Then*

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = 0 \quad a.s.$$

*Proof.* By Theorem 7.2, $\sum_n X_n/n$ converges a.s.

By Kronecker's Lemma, $(X_1 + \cdots + X_n)/n \to 0$ a.s. □

7.3. **SLLN for iid case.**

**Theorem 7.5** (Kolmogorov)**.** *Let $(X_n)_{n\geq 1}$ be iid with zero mean. Then*

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = 0 \quad a.s.$$

*If $\mathbb{E}(|X_1|) = \infty$ then*

$$\limsup_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \infty \quad a.s.$$

*Proof.* In order to apply previous results we need a moment restriction. For this, define

$$Y_n = \begin{cases} X_n & \text{if } |X_n| \leq n \\ 0 & \text{if } |X_n| > n \end{cases}, \qquad X_n = Y_n + Z_n.$$

We next show that a.s., $Z_n = 0$ for all sufficiently large $n$.

By Borel-Cantelli it is sufficient to show that $\sum_n P(Z_n \neq 0) < \infty$, or equivalently that $\sum_n P(|X_n| > n) < \infty$. For this let $E_n = \{x : |x| > n\} = [-n, n]^c$ and let $P$ denote the probability distribution for $X_1$. Then

$$\begin{aligned}\sum_{n\geq 1} P(|X_n| > n) &= \sum_{n\geq 1} \int \mathbb{I}_{E_n}(x)\, dP(x) \\ &= \int \sum_{n\geq 1} \mathbb{I}_{E_n}(x)\, dP(x) \\ &\leq \int |x|\, dP(x) = \mathbb{E}(X_1) < \infty.\end{aligned}$$

In order to apply Theorem 7.4 to $Y_n$ first note

$$\mathbb{E}(Y_n - \mathbb{E}(Y_n))^2 \leq \mathbb{E}(Y_n)^2 = \int_{[-n,n]} x^2\, dP(x).$$

It follows that

$$\begin{aligned}
\sum_{n\geq 1}\frac{\operatorname{var}(Y_n)}{n^2} &\leq \sum_{n\geq 1}\frac{1}{n^2}\int_{[-n,n]} x^2\,dP(x)\\
&= \sum_{n\geq 1}\sum_{1\leq \ell\leq n}\frac{1}{n^2}\int \mathbb{I}_{F_\ell}(x)x^2\,dP(x)\quad \text{where } F_\ell := \{x : \ell-1<|x|\leq \ell\}\\
&= \sum_{\ell\geq 1}\sum_{n\geq \ell}\frac{1}{n^2}\int \mathbb{I}_{F_\ell}(x)x^2\,dP(x)\\
&\leq \sum_{\ell\geq 1}\sum_{n\geq \ell}\frac{\ell}{n^2}\int \mathbb{I}_{F_\ell}(x)|x|\,dP(x).
\end{aligned}$$

But

$$\sum_{n\geq \ell}\frac{1}{n^2}\sim\int_\ell^\infty\frac{dt}{t^2}\leq\frac{c}{\ell},$$

so

$$\sum_{n\geq 1}\frac{\operatorname{var}(Y_n)}{n^2}\leq c\sum_{\ell\geq 1}\int \mathbb{I}_{F_\ell}(x)|x|\,dP(x) = c\int |x|\,dP(x) = c\mathbb{E}(X_1)<\infty.$$

From Theorem 7.4 with $\mu_n = \mathbb{E}(Y_n)$,

$$\frac{Y_1+\cdots+Y_n}{n}-\frac{\mu_1+\cdots+\mu_n}{n}\to 0\quad\text{a.s.}$$

But

$$\mu_n = \int_{[-n,n]}|x|\,dP(x)\to\mathbb{E}(X_1)=0\text{ as } n\to\infty.$$

Hence

$$\frac{Y_1+\cdots+Y_n}{n}\to 0\quad\text{a.s.}$$

Since eventually $Z_n = 0$ a.s., it follows that

$$\frac{Z_1+\cdots+Z_n}{n}\to 0\quad\text{a.s.}$$

Thus

$$\frac{X_1+\cdots+X_n}{n}\to 0\quad\text{a.s.}$$

This completes the proof of the main part of the theorem.

For the last part, assume $\mathbb{E}(|X_1|)=\infty$.
Suppose $C>0$ and let

$$\begin{aligned}
A_n &= \{\omega : |X_n|\geq Cn\}\subset\Omega\\
E_n &= \{x : |x|\geq Cn\}\subset\mathbb{R}.
\end{aligned}$$

Then $P(A_n) = P(E_n)$, where the second $P$ is the probability on $\mathbb{R}$ induced by $X_n$, and is independent of $n$.

Hence

$$\begin{aligned}
\sum_n P(A_n) &= \sum_n P(E_n)\\
&= \sum_n\int \mathbb{I}_{E_n}(x)\,dP(x)\\
&= \int\sum_n \mathbb{I}_{E_n}(x)\,dP(x)\\
&\sim c^{-1}\int |x|\,dP(x)=\infty.
\end{aligned}$$

Since the $A_n$ are independent, by Borel-Cantelli $P(A_n \text{ i.o.}) = 1$. That is, almost surely

$$\frac{|X_n|}{n} \geq C \text{ i.o.}$$

Assume

$$\limsup_{n\to\infty} \frac{|X_1 + \cdots + X_n|}{n} < \infty$$

on a set of $\omega$ of positive measure. Then for some $K > 0$, there exists $A \subset \Omega$ with positive measure such that for $\omega \in A$ and for $n \geq n_0(\omega)$,

$$-K \leq \frac{X_1 + \cdots + X_n}{n}, \frac{X_1 + \cdots + X_{n-1}}{n} \leq K,$$

and so by subtraction if $\omega \in A$ and $n \geq n_0$,

$$\frac{|X_n|}{n} \leq 2K.$$

Taking $C = 3K$ gives a contradiction. □

## 8. Renewal Theorem

We follow [Fal97, Chapter 7].

**Definition 8.1.** Suppose $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable and $\mu$ is a Borel probability measure on $[0, \infty)$. Then the corresponding *renewal equation* is

$$\begin{aligned} f(t) &= g(t) + \int_0^\infty f(t-y)\, d\mu(y) \quad t \in \mathbb{R} \\ \text{i.e. } f &= g + f * \mu \\ \text{or } f(t) &= g(t) + \mathbb{E} f(t - X) \text{ where } \operatorname{dist} X = \mu. \end{aligned} \tag{13}$$

*Remark* 8.2. Think of $t$ s time. Then the integral in (13) is a weighted average of $f$ at times earlier than $t$ (and at $t$ if $\mu$ has an atom at 0). Moreover, $g(t)$ can be thought of as an error between $f(t)$ and this integral. □

*Remark* 8.3. Consider the *renewal process* defined by

$$T_0 = 0, \quad T_n = X_1 + \cdots + X_n \text{ if } n \geq 1. \tag{14}$$

where $X_n \geq 0$ (usually $> 0$) are iid with distribution $\mu$.

Note that

$$\mu[0, t] = P\{X_1 \leq t\}, \quad \mu^{*n}[0, t] = P\{X_1 + \cdots + X_n \leq t\} = P\{T_n \leq t\}. \tag{15}$$

(This corresponds to installing a light bulb at $t = 0$, and subsequently immediately upon failure of the previous bulb. Then the $T_n$ are the installation or renewal times.)

The associated *renewal counting process* $(N_t)_{t\geq 0}$ is the number of renewals up to and including time $t$. That is

$$N_t = \operatorname{card}\{n : T_n \leq t\} = \sum_{n\geq 0} \mathbb{I}_{\{T_n \leq t\}}. \tag{16}$$

□

*Remark* 8.4. The *renewal function* and *renewal measure* (both denoted $U$) are defined by

$$U(t) = U[0, t] = \mathbb{E}\, N(t).$$

This is the expected number of renewals up to time $t$, and $U(A)$ is the expected number of renewals in $A$. Moreover,

$$U[0, t] = \sum_{n\geq 0} \mathbb{E}\, \mathbb{I}_{\{T_n \leq t\}} = \sum_{n\geq 0} P\{T_n \leq t\} = \mu^{*n}[0, t].$$

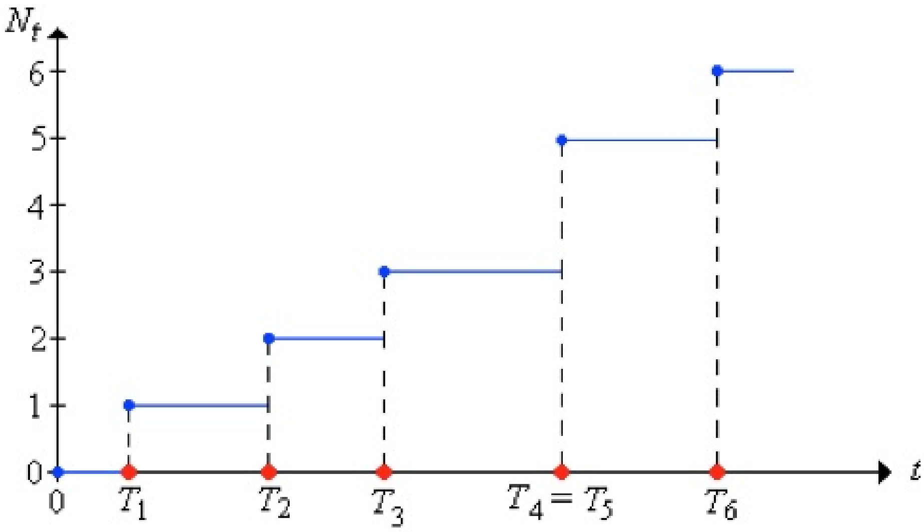


FIGURE 1. Renewal process (from [Che09])

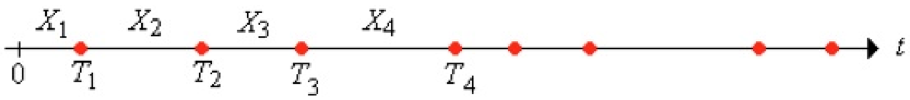


FIGURE 2. Renewal process (from [Che09])

That is

$$U = \sum_{n \geq 0} \mu^{*n}, \quad \text{where } \mu^{*0} = \delta_0. \tag{17}$$

Note $T_0 = 0$. This also follows from Proposition 8.6.

For $t \geq 0$,

$$\begin{aligned} U(t) &= \mathbb{E}_y \mathbb{E}(N(t) \mid X_1 = y) \\ &= \mathbb{E}_y (1 + U(t-y)) \quad \text{(since the process starts again at } X_1) \\ &= 1 + \int U(t-y)\, d\mu(y). \end{aligned}$$

That is, $U$ satisfies the renewal equation (13) with $g = \mathcal{X}_{[0,\infty)}$.

It follows from the renewal theorem that in the non-arithmentic case, *$U$ is approximately a multiple of Lebesgue measure $L^1$*. More precisely,

*Proposition* 8.5. *If $\mu$ is non-arithmetic, then*

$$U[t, t+h] \to \lambda^{-1} h \text{ as } t \to \infty,$$

*for every $h > 0$. Moreover, if $\mu$ is $\tau$-arithmetic then the same is true provided $h$ is a multiple of $\tau$.*

*Proof.* Let $g = \mathcal{X}_[0, h]$ in the Renewal Theorem 8.10. □

The renewal equation, under quite general conditions, has a unique solution given by an infinite series expansion.

More precisely, we make the following assumptions *** these are only needed for the renewal theorem. The following proposition is true under weaker hypotheses, sufficient to include the case with $U(t)$ and $g(t)$ as above ***

(1) $\lambda := \mathbb{E}(X) = \int_0^\infty t\, d\mu(t) < \infty$, where $\operatorname{dist} X = \mu$;
(2) $\mu$ is not concentrated at 0;
(3) $|g(t)| \le ce^{-\alpha|t|}$ for some $\alpha > 0$, and $g$ has a discrete set of discontinuities (more generally, $g$ is "directly Riemann integrable").

Note that condition (3) on $g$ is *not* satisfied in in Example 8.3.

Let $\mathcal{F}$ be the set of Borel measurable $f : \mathbb{R} \to \mathbb{R}$ such that $f(t) \to 0$ as $t \to -\infty$ and $f$ is bounded on each $(-\infty, a]$.

**Proposition 8.6** (Solution of the renewal equation)**.** *Under the previous hypotheses *** and more generally as noted before *** there is a unique solution of the renewal equation given by*

$$
\begin{aligned}
f &= \sum_{n\ge 0} g * \mu^{*n} = g * U, \\
\text{i.e. } f(t) &= \sum_{n\ge 0} \mathbb{E}\, g(t - (X_1 + \cdots + X_n)) \qquad (18)\\
&= \sum_{n\ge 0} \int_0^\infty \cdots \int_0^\infty g(t - y_1 - \cdots - y_n)\, d\mu(y_1)\ldots d\mu(y_n).
\end{aligned}
$$

*Moreover, $f$ is bounded, and if $g$ is continuous then $f$ is uniformly continuous.*

*Proof.* The formal idea is that if $f(t) = \sum_{n\ge 0} (g * \mu^{*n})(t)$ then

$$
\begin{aligned}
f(t) &= g(t) + \sum_{n\ge 0} ((g * \mu^{*n}) * \mu)(t) \\
&= g(t) + \int \sum_{n\ge 0} (g * \mu^{*n})(t - y)\, d\mu(y) \\
&= g(t) + \int f(t - y)\, d\mu(y).
\end{aligned}
$$

The justification for the various steps and for the regularity results follow from the hypotheses. □

For the Renewal Theorem 8.10 we will need to consider two cases for $\mu$.

**Definition 8.7.** The measure $\mu$ is *$\tau$-arithmetic* if

$$E := \operatorname{spt}\mu \subset \{a + \tau k : k \in \mathbb{N}\}$$

for some $a \in \mathbb{R}$ (take $a \in [0, \tau)$ w.l.o.g.), and $\tau$ is the greatest such positive number.

Otherwise, $\mu$ is *non-arithmetic*.

*Remark* 8.8. If $\mu$ is $\tau$-arithmetic and $f$ satisfies the renewal equation, then it is clear from the first form of (13) that, for each *fixed* $t \in \mathbb{R}$, this gives a relationship involving only

$$f(t - k\tau),\ g(t),\ \mu\{k\tau\}, \quad \text{for } k \in \mathbb{N}_0.$$

Similarly, if $f$ satisfies the renewal equation, then it is clear from the second form of (18) that, for each *fixed* $t \in \mathbb{R}$, this gives a relationship involving only

$$f(t),\ g(t - k\tau),\ \mu\{k\tau\}, \quad \text{for } k \in \mathbb{N}_0.$$

*Remark* 8.9. The renewal theorem below refers to the limit behaviour of the solution $f(t)$ to the renewal equation as $t \to \infty$. Recall $f(t) = \sum_{k\ge 0} \mathbb{E}\, g(t - (X_1 + \cdots + X_n))$.

Since $f = g * U$ and $U$ is asymptotically $L^1$ normalised by the mean of $\mu$, (at least in the non-arithmetic case) we expect that asymptotically $f$ is the integral of $g$ normalised by the mean of $\mu$. More carefully, we argue as follows.

*Non-arithmetic case.* In order to find an approximation to $f(t)$ for large $t$, approximate $g$ by a sum of functions of the form $\alpha\mathcal{X}_{[a,b]}$. First suppose $g$ is itself a summand of this form, $t >> b$ and $b - a << \lambda$. The probability that $t - (X_1 + \cdots + X_n) \in [a, b]$ for some $n$ is approximately $[b-a]/\lambda$, and this is the same for the probability that $t-(X_1+\cdots+X_n) \in [a, b]$ for exactly one $n$. It follows that $f(t) \approx \alpha[b - a]/\lambda$. By summing, for general $g$ it follows that $f(t) \approx \lambda^{-1} \int g(x)\, dx$.

*Arithmetic case.* In order to find an approximation to $f(t)$ for large $t$, it follows from Remark 8.8 that the relevant arguments for $g$ are $t-k\tau$ for $k \in \mathbb{N}_0$. First suppose $g(t-k'\tau) = \alpha$ for some $k'$ and that otherwise $g(t - k\tau) = 0$, and suppose $t >> t - k'\tau$. The probability that $t - (X_1 + \cdots + X_n) = t - k'\tau$ for some (and hence exactly one) $n$ is approximately $1/\lambda d$. It follows that $f(t) \approx \lambda^{-1}\alpha$. By summing, for general $g$, it follows that $f(t) \approx \lambda^{-1} \sum_{j=-\infty}^{\infty} g(t + j\tau)$. □

**Theorem 8.10** (Renewal theorem)**.** *Suppose the hypotheses as for Proposition 8.6 and that $f \in \mathcal{F}$ is the solution of the renewal theorem. If $\mu$ is non-arithmetic then*

$$\lim_{x\to\infty} f(x) = \lambda^{-1} \int g(t)\, dt. \tag{19}$$

*If $\mu$ is $\tau$-arithmetic then for $t_0 \in [0, \tau)$,*

$$\lim_{k\to\infty} f(t_0 + k\tau) = \lambda^{-1} \sum_{j=-\infty}^{\infty} g(t_0 + j\tau). \tag{20}$$

*Proof.* **Method 1.** This makes rigorous the informal argument in Remark 8.9 by using acoupling argument.

**Method 2.** We outline the non arithmetic case.

*Step (a).* From (13) one gets

$$\int_{-\infty}^{x} g(t)\, dt = \int f(t)\, \psi(x - t)\, dt = (f * \psi)(x) \quad \text{where } \psi(t) = \begin{cases} 0 & t < 0 \\ \mu[t, \infty) & t \geq 0 \end{cases}.$$

*Step (b).* Hence

$$(f * \psi)(x) \to \int g(t)\, dt \quad \text{as } x \to \infty.$$

(Think of $(f * \psi)(x)$ as a weighted average of values of $f$ at points $y < x$.)

*Step (c).* By Wiener's theorem, since we can show $\widehat{\psi}(u) \neq 0$,

$$(f * \phi)(x) \to \frac{\int \phi}{\int \psi} \int g(t)\, dt \quad \text{as } x \to \infty,$$

for *any* $\phi \in L^1(\mathbb{R})$.

We can check that $\int \psi = \lambda$.

*Step (d).* Setting $\phi$ to be an approximation to the Dirac $\delta$-function and using the uniform continuity of $f$, it follows that

$$f(x) \to \lambda^{-1} \int g(t)\, dt \quad \text{as } x \to \infty.$$

*Step (e).* If $g$ has a discrete set of discontinuities, then we approximate $g$ by continuous functions.

(Arithmetic case?) □

We use the following, which is just a restatement of the previous theorem for $\mu$ with finite discrete support.

**Corollary 8.11.** *Suppose $m \geq 2$, $t_1, \dots, t_m > 0$ are "times", and $p_1, \dots, p_m$ are probabilities, so that $\sum_i p_i = 1$. Let $g$ be as before Proposition 8.6 and let $f$ satisfy the renewal equation*

$$f(t) = g(t) + \sum_i p_i f(t - t)i). \tag{21}$$

*Let $\lambda = \sum_i p_i t_i$.*

*If $\{t_1, \dots, t_m\}$ is non-arithmetic then*

$$\lim_{t\to\infty} f(t) = \lambda^{-1} \int g(t)\, dt.$$

*If $\{t_1, \dots, t_m\}$ is $\tau$-arithmetic then*

$$\lim_{k\to\infty} f(t_0 + k\tau) = \lambda^{-1} \sum_{k=-\infty}^{\infty} g(t_0 + k\tau),$$

*for all $t_0 \in [0, \tau)$.*

## 9. CJM Processes

9.1. **Notation and Non-probabilistic aspects.** Consider a population of individuals with an *initial ancestor* denoted by $\phi$. This individual will have a finite number of children, each of these will have a finite number of children, etc. Each individual apart from the initial ancestor has exactly one parent and there is no notion of breeding in this model.

(Later we will impose a notion of absolute time, and of birth and death times.)

The set of all individuals (alive or dead) is naturally represented by a *tree* $T \subset \bigcup_{k\geq 0} \mathbb{N}^k$, where $\mathbb{N}^0 = \{\emptyset\}$ and $\mathbb{N}^k$ is the set of finite sequences $\boldsymbol{i} = i_1 \dots i_k$ of positive integers. We use the standard notations $|\boldsymbol{i}|$ for the length of $\boldsymbol{i}$, $\boldsymbol{i}|_k$ for truncation and $\boldsymbol{ij}$ for concatentation.

Motivated by the above we require:

(1) $\emptyset \in T$;
(2) $\boldsymbol{i} \in T$ implies (the unique $k$th generation ancestor) $\boldsymbol{i}|_k \in T$ for $k < |\boldsymbol{i}|$;
(3) $\boldsymbol{i}1, \dots, \boldsymbol{i}N^{\boldsymbol{i}} \in T$ and $\boldsymbol{i}(N^{\boldsymbol{i}}+1), \boldsymbol{i}(N^{\boldsymbol{i}}+2), \dots \notin T$, where $N^{\boldsymbol{i}}$ is the *number of children of* $\boldsymbol{i}$.

Associated with each $\boldsymbol{i} \in T$ is a *life-story* $U^{\boldsymbol{i}} = (L^{\boldsymbol{i}}, \xi^{\boldsymbol{i}})$ where

(1) $L^{\boldsymbol{i}} \in [0, \infty)$ is the *lifetime* of $\boldsymbol{i}$;
(2) $\xi^{\boldsymbol{i}} : [0, \infty) \to \mathbb{N}_0$[1] is a bounded non-decreasing right-continuous function, and $\xi^{\boldsymbol{i}}(t)$ is the number of births from $\boldsymbol{i}$ up to and including time $t$. We *assume* $\xi(0) = 0$.

The jumps in $\xi^{\boldsymbol{i}}$ determine age $t^{\boldsymbol{i}}(j)$ of $\boldsymbol{i}$ at the birth of the $j$th child of $\boldsymbol{i}$. The number of births at age $t$ is the size of the jump at $t$. The total number of births for $\boldsymbol{i}$ is denoted $N^{\boldsymbol{i}}$. The ages at time of berth satisfy

$$0 < t^{\boldsymbol{i}}(1) \leq t^{\boldsymbol{i}}(2) \leq \cdots \leq t^{\boldsymbol{i}}(N^{\boldsymbol{i}}) \leq L^{\boldsymbol{i}}. \tag{22}$$

More precisely, *define*

$$N^{\boldsymbol{i}} := \xi[0, \infty), \quad t^{\boldsymbol{i}}(j) := \inf\{t : \xi^{\boldsymbol{i}}(t) \geq j\}. \tag{23}$$

We also w.l.o.g. make the assumption

$$t^{\boldsymbol{i}}(N^{\boldsymbol{i}}) \leq L^{\boldsymbol{i}}. \tag{24}$$

Then (22) follows, as does

$$\xi^{\boldsymbol{i}}(t) = \max\{j : t^{\boldsymbol{i}}(j) \leq t\}. \tag{25}$$

See Figure 1, where a different notation is used.

[1] $\mathbb{N}_0$ is the set of natural numbers together with 0.

In the usual way, $\xi^{\boldsymbol{i}}$ is the distribution function for a measure, also denoted by $\xi^{\boldsymbol{i}}$. Thus

$$\xi^{\boldsymbol{i}}(s,t] = \xi^{\boldsymbol{i}}(t) - \xi^{\boldsymbol{i}}(s), \quad \xi^{\boldsymbol{i}}(t) = \xi^{\boldsymbol{i}}(0,t].$$

The measure $\xi^{\boldsymbol{i}}$ is a sum of Dirac measures, one for each birth and counted with multiplicities. More precisely,

$$\xi^{\boldsymbol{i}} = \sum_{j=1}^{N^{\boldsymbol{i}}} \delta_{t^{\boldsymbol{i}}(j)}. \tag{26}$$

The *time at the birth* of an individual is defined recursively by

$$\sigma_\emptyset = 0, \quad \sigma_{\boldsymbol{i}j} = \sigma_{\boldsymbol{i}} + t^{\boldsymbol{i}}(j). \tag{27}$$

Thus if we attach the time $t^{\boldsymbol{i}}(j)$ to the edge from $\boldsymbol{j}$ to $\boldsymbol{j}k$ then the time at birth of $\boldsymbol{i}$ is the sum of the times along all edges in $i_1 \dots i_{|\boldsymbol{i}|}$. That is

$$\sigma_{\boldsymbol{i}} = t^{\emptyset}(i_1) + t^{i_1}(i_2) + \cdots + t^{i_1\dots i_{n-1}}(i_n) = \sum_{k=1}^{n} t^{i_1\dots i_{k-1}}(i_k) \quad \text{if } |\boldsymbol{i}| = n \geq 1. \tag{28}$$

*Example* 9.1. We use standard notation.

Consider a fractal set indexed by a tree $T$ in the usual manner. That is, the $n$-cell $\Delta_{\boldsymbol{i}}$ is replaced by a scaled copy of $F^{\boldsymbol{i}}(G_0) = \bigcup_{j=1}^{M} f_j^{\boldsymbol{i}}(G_0)$, where $F^{\boldsymbol{i}} = \{f_1^{\boldsymbol{i}}, \dots, f_M^{\boldsymbol{i}}\}$ is an IFS of similarities in $\mathbb{R}^k$ with contraction ratios $\ell_1^{\boldsymbol{i}} \geq \cdots \geq \ell_M^{\boldsymbol{i}}$.

In this case the ages of $\boldsymbol{i}$ at times of giving birth and at death are defined by

$$t^{\boldsymbol{i}}(j) = \log 1/\ell_j^{\boldsymbol{i}}, \quad L^{\boldsymbol{i}} = \log 1/\ell_M^{\boldsymbol{i}}.$$

In particular, from (28) by setting

$$t^{\boldsymbol{i}}(j) = \log\left(\ell_j^{\boldsymbol{i}}\right)^{-1}, \quad \ell_j^{\boldsymbol{i}} = \exp\left(-t^{\boldsymbol{i}}(j)\right),$$

it follows that the time at birth of the cell $\boldsymbol{i}$ is

$$\sigma_{\boldsymbol{i}} = \log \ell_{\boldsymbol{i}}^{-1}, \quad \text{i.e. } \ell_{\boldsymbol{i}} = \exp(-\sigma_{\boldsymbol{i}}), \tag{29}$$

where as usual $\ell_{\boldsymbol{i}} := \prod_{k=1}^{n} \ell_{i_k}^{i_1\dots i_{k-1}}$ if $|\boldsymbol{i}| = n$.

Note for future reference that if $F := \{f_1, \dots f_m\}$ and $t_j := \log \ell_j^{-1}$ then

$$\sum_j \ell_j^\alpha = 1 \iff \sum_j e^{-\alpha t_j} \left(\text{i.e. } \int_0^\infty e^{-\alpha t} \xi(dt)\right) = 1, \tag{30}$$

and more generally

$$\sum_{\boldsymbol{i}\in\Lambda} \ell_{\boldsymbol{i}}^\alpha = 1 \iff \sum_{\boldsymbol{i}\in\Lambda} e^{-\alpha\sigma_{\boldsymbol{i}}} = 1. \tag{31}$$

## References


[Ash00] Robert B. Ash, *Probability and measure theory*, 2nd ed., Harcourt/Academic Press, Burlington, MA, 2000. With contributions by Catherine Doléans-Dade.

[Che09] Kani Chen, *Advanced probability Math 541*, 2009. online notes.

[Fal97] Kenneth Falconer, *Techniques in fractal geometry*, John Wiley & Sons Ltd., 1997.

[Fel71] William Feller, *An Introduction to probability Theory and its Applications*, 2nd ed., Wiley, 1971.

[Lam66] John Lamperti, *Probability. A survey of the mathematical theory*, W. A. Benjamin, Inc., New York-Amsterdam, 1966.