Section 1: Course Introduction and the Law of Large Numbers

Contents

1.1 Introduction

Emphasis on:

- key model classes that lend themselves to tractability
- key tools used to render these models tractable

What makes a model *tractable* then? Usually, by "tractability", we are seeking one of the following results:

- analytical "closed forms" (notion of closed form is "in the eyes of the beholder")
- approximations
  - this was the primary approach used to simplify model analysis historically
  - mathematical justification based on "limit theorems"
  - "limit theorems" rely on precise notions of convergence
- numerics
  - sampling-based (Monte Carlo)
  - non-sampling based (e.g. solving systems of linear equations, ordinary differential equations (ODE's), partial differential equations (PDE's))

Model formulation should rest, in part, on questions of tractability (e.g. Can we easily compute numerically for the given model?)

Primary focus in this course is on *Markov* models.

1.2 Taxonomy of Markov Models

Discrete-time vs continuous-time
Discrete state space vs non-discrete state space

A *stochastic process* is a mapping $X : T \times \Omega \to \mathcal{S}$; T is the time-parameter (index) set, Ω is the underlying sample space, and S is the "state space" in which X takes values.

Discrete-time: $T = \mathbb{Z}^+$ or $\mathbb{Z}$. In this case, $X = (X_n : n \geq 0)$ or $X = (X_n : n \in \mathbb{Z})$; X is often referred to as a *stochastic sequence*. When X is Markov, such stochastic sequences are called Markov chains.

Continuous-time: $T = \mathbb{R}^+$ or $\mathbb{R}$. Here, $X = (X(t) : t \geq 0)$ or $X = (X(t) : t \in \mathbb{R})$; X is called a *process*.

In modern terminology, it is common to reserve the term "process" for continuous time.

Discrete state space: $\mathcal{S}$ is either finite or countably infinite.

Continuous state space: $\mathcal{S}$ is some "nice" subset of $\mathbb{R}^d$,

General state space: $\mathcal{S}$ can be completely general (so this subsumes both discrete state space and continuous state space.)

Discrete-time Markov chains
Continuous-time Markov jump processes
Continuous-time Markov diffusion processes (also known as solutions to "stochastic differential equations")

1.3 Introduction to Approximations and Limit Theorems

A fundamental building block of much of the theory is "random walk".

Definition 1.3.1 A sequence $S = (S_n : n \geq 0)$ is called a *random walk* if

$$S_n = Z_1 + Z_2 + \cdots + Z_n$$

where the Z_i's are iid $\mathbb{R}^d$-valued random vectors.

An important special case is where each $Z_i \stackrel{\mathcal{D}}{=} Ber(p)$. In this case, $S_n \stackrel{\mathcal{D}}{=} Bin(n, p)$, i.e.

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Remark 1.3.1 Note that S_n might represent the total number of defectives in the first n items inspected on an assembly line.

Because of the presence of the combinatorial coefficients, we seek simpler approximations.

1.3.1 Poisson Approximation

When n is large and p is small, the following approximation is often very good:

$$Bin(n,p) \stackrel{\mathcal{D}}{\approx} Poisson(np).$$

This can be used to conclude, for example, that the number of typos in a paper should be approximately Poisson distributed. (Why?)

Mathematical Justification for Poisson Approximation

For each $k \geq 0$,

$$P(Bin(n,p) = k) \to P(Poisson(\lambda) = k)$$

as $n \to \infty$, provided that $n \to \infty$ and $np \to \lambda \in (0, \infty)$. (Prove this.)

Remark 1.3.2 This limit result is asserting that

$$Bin(n,p) \Rightarrow Poisson(\lambda)$$

as $n \to \infty$, provided that $n \to \infty$ and $np \to \lambda \in (0, \infty)$, where $\Rightarrow$ denotes *weak convergence* (and is also known as "convergence in distribution").

Remark 1.3.3 The above Poisson approximation is a special case of the following approximation. Suppose that $Z_i \stackrel{\mathcal{D}}{=} Ber(p_i)$ for all $i \geq 1$, where the p_i's are small. Then,

$$S_n \stackrel{\mathcal{D}}{\approx} Poisson\left(\sum_{i=1}^{n} p_i\right)$$

This approximation is useful in many settings (e.g. Suppose Z_i is 1 or 0 depending on whether firm i defaults or not; then S_n is the total number of defaults in a portfolio of n firms.)

1.3.2 Normal Approximation

When n is large (and p is not too close to either 0 or 1), the normal approximation can be very effective:

$$Bin(n,p) \stackrel{\mathcal{D}}{\approx} np + \sqrt{np(1-p)}N(0,1),$$

where $N(0,1)$ is a normal random variable with mean 0 and variance 1.

Mathematical Justification for Normal Approximation

For each $x \in \mathbb{R}$,

$$P\left(\frac{B(n,p) - np}{\sqrt{np(1-p)}} \leq x\right) \to P(N(0,1) \leq x)$$

as $n \to \infty$, provided $0 < p < 1$.

Remark 1.3.4 This limit result is asserting that

$$\frac{B(n,p) - np}{\sqrt{np(1-p)}} \Rightarrow N(0,1)$$

as $n \to \infty$, provided that $0 < p < 1$. This is a special case of the *central limit theorem* (CLT). A more general version states that if $0 < \operatorname{var} Z_i < \infty$, then

$$\frac{S_n - nEZ_1}{\sqrt{n \operatorname{var} Z_1}} \Rightarrow N(0,1)$$

as $n \to \infty$. This suggests the approximation

$$S_n \stackrel{\mathcal{D}}{\approx} EZ_1 + \sqrt{n \operatorname{var} Z_1} N(0,1)$$

when n is large. When do we expect the approximation

$$P(S_n \leq x) \approx P(nEZ_1 + \sqrt{n \operatorname{var} Z_1} N(0,1) \leq x)$$

to be good?

- when n is large
- when x is within a few "standard derivation" of $nEZ_1 = ES_n$.

Limit theorems give us information on when we can expect approximations to be good!

1.4 Weak Convergence

The general notion of weak convergence of $(X_n : n \geq 0)$ to X_∞, written as $X_n \Rightarrow X_\infty$ as $n \to \infty$, must subsume the two above special case, but should also permit us to assert that

$$X_n = \frac{1}{n} \Rightarrow 0 = X_\infty$$

as $n \to \infty$. But note that

$$P(X_n \leq x) \to P(X_\infty \leq x)$$

as $n \to \infty$ when $x \neq 0$. On the other hand,

$$0 = P(X_n \leq 0) \nrightarrow P(X_\infty \leq 0) = 1.$$

So, $P(X_n \leq 0)$ does not converge to the "correct limit" at the discontinuity point $x = 0$. As a result, the general version of weak convergence of real-valued random variables reads as follows:

Definition 1.4.1 We say that $X_n \Rightarrow X_\infty$ as $n \to \infty$ if

$$P(X_n \leq x) \to P(X_\infty \leq x)$$

as $n \to \infty$ at every point x which is a continuity point of $P(X_\infty \leq \cdot)$.

1.5 The Monte Carlo Method

A powerful numerical tool for simulating stochastic models is to use sampling-based methods (also known as "Monte Carlo" methods).

Goal: Compute EZ

Method:

1. Select the sample size (ie. number of computer experiments/simulations n).
2. Generate n iid copies $Z_1, Z_2, \ldots, Z_n$.
3. Estimate EZ via $\bar{Z}_n \triangleq \frac{1}{n}\sum_{i=1}^{n} Z_i$

When $E|Z| < \infty$, the *(weak) law of large numbers* (WLLN) asserts that

$$\bar{Z}_n \Rightarrow EZ \tag{1.5.1}$$

as $n \to \infty$.

Remark 1.5.1 The above limit result is asserting that $(\bar{Z}_n : n \geq 1)$ is a *consistent estimator* for EZ.

Remark 1.5.2 A sequence $(X_n : n \geq 1)$ is said to *converge in probability* if there exists X_∞ such that for any $\epsilon > 0$,

$$P(|X_n - X_\infty| > \epsilon) \to 0$$

as $n \to \infty$. When $(X_n : n \geq 1)$ converges to X_∞ in probability, we write $X_n \xrightarrow{p} X_\infty$ as $n \to \infty$.

Problem 1.5.1 Prove that if $X_n \xrightarrow{p} X_\infty$ as $n \to \infty$, then $X_n \Rightarrow X_\infty$ as $n \to \infty$.

Problem 1.5.2 If X_∞ is deterministic, prove that $X_n \Rightarrow X_\infty$ implies that $X_n \xrightarrow{p} X_\infty$ as $n \to \infty$. (Hence, weak convergence and convergence in probability are equivalent when X_∞ is deterministic.)

Remark 1.5.3 The weak law is equivalent to asserting that

$$\bar{Z}_n \xrightarrow{p} EZ$$

as $n \to \infty$.

The question arises as to how fast $\bar{Z}_n$ converges to EZ. This is answered by the central limit (normal) approximation:

$$\bar{Z}_n \overset{\mathcal{D}}{\approx} EZ + \frac{\sigma}{\sqrt{n}} N(0,1)$$

when n is very large.

This implies that:

- the error decreases to zero at "square root rate" $n^{-\frac{1}{2}}$.

- the error is approximately normally distributed.
- the rate of convergence (to first order) depends on the "problem instance" only through a single problem parameter, namely $\sigma = \sqrt{\operatorname{var} Z}$.

Given the slow rate of convergence, good error diagnostics or the Monte Carlo method are important. This is typically accomplished via *confidence intervals*.

1.6 Confidence Intervals for the Monte Carlo Method

To provide an approximate $100(1-\delta)\%$ (confidence interval) for EZ, we compute the *sample standard deviation* s_n given by

$$s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2},$$

in addition to the *sample mean* $\bar{Z}_n$. Having computed a value of z for which

$$P(-z \le N(0,1) \le z) = 1 - \delta,$$

the approximate $100(1-\delta)\%$ confidence interval for EZ is given by

$$\left[\bar{Z}_n - z\frac{s_n}{\sqrt{n}}, \bar{s}_n + z\frac{x_n}{\sqrt{n}}\right].$$

The justification for this random interval as a confidence interval rests upon the following result:

Proposition 1.6.1 If $0 < \operatorname{var} Z < \infty$, then

$$P\left(EZ \in \left[\bar{Z}_n - z\frac{s_n}{\sqrt{n}}, \bar{Z}_n + z\frac{s_n}{\sqrt{n}}\right]\right) \to 1 - \delta$$

as $n \to \infty$.

How does one (rigorously) prove this?

- Step 1. The CLT asserts that

  $$\frac{S_n - nEZ}{\sqrt{n}\sigma} = n^{-1/2}\frac{(\bar{Z}_n - EZ)}{\sigma} \Rightarrow N(0,1)$$

  as $n \to \infty$.
- Step 2. Prove that

  $$s_n \xrightarrow{p} \sigma$$

  as $n \to \infty$.

  Remark 1.6.1 This starts from observing that the WLLN guarantees that

  $$1/n\sum_{i=1}^{n} Z_i^2 \xrightarrow{p} EZ^2$$

and

$$\bar{Z}_n \xrightarrow{p} EZ$$

as $n \to \infty$. It follows that

$$(1/n \sum_{i=1}^{n} Z_i^2, \bar{Z}_n) \xrightarrow{p} (EZ^2, EZ)$$

as $n \to \infty$. (Prove this!) Furthermore, if $\beta_n \xrightarrow{p} \beta_\infty$ as $n \to \infty$, it is evident that $h(\beta_n) \xrightarrow{p} h(\beta_\infty)$ as $n \to \infty$, provided h is continuous. (Prove this!) Setting $h(x_1, x_2) = \sqrt{x_1^2 - x_2^2}$, we conclude that $S_n \xrightarrow{p} \sigma$ as $n \to \infty$.

- Step 3. Prove that

$$n^{1/2} \frac{(\bar{Z}_n - EZ)}{s_n} \Rightarrow N(0,1)$$

as $n \to \infty$.

Remark 1.6.2 Clearly, $\sigma/s_n \xrightarrow{p} 1$ as $n \to \infty$. We would like to conclude that

$$\begin{aligned} n^{1/2} \frac{(\bar{Z}_n - EZ)}{s_n} &= n^{1/2} \frac{(\bar{Z}_n - EZ)}{\sigma} \cdot \frac{\sigma}{s_n} \\ &\Rightarrow N(0,1) \cdot 1 \\ &= N(0,1) \end{aligned}$$

as $n \to \infty$. But we need to be careful. It is *not* true that if $X_n \Rightarrow X_\infty$ and $\beta_n \Rightarrow \beta_\infty$, then $X_n \beta_n \Rightarrow X_\infty \beta_\infty$. It is not even true that if $X_n \xrightarrow{p} X_\infty$ and $\beta_n \xrightarrow{p} \beta_\infty$, then $X_n \beta_n \xrightarrow{p} X_\infty \beta_\infty$ as $n \to \infty$. (This basic problem is asserting that the distribution of $X_n \beta_n$ clearly depends on the joint distribution of X_n and β_n. But the hypotheses give us no information on the joint distribution.) However, the following line of argument is correct. Suppose that $X_n \Rightarrow X_\infty$ and $\beta_n \Rightarrow c$ where c is deterministic. Then $(X_n, \beta_n) \Rightarrow (X_\infty, c)$ as $n \to \infty$. (Prove this!)

We also need to know the continuous mapping principle and its extended version.

Continuous Mapping Principle: Suppose $(X_n : n \geq 1)$ is a sequence of $\mathbb{R}^d$-valued random variables for which $X_n \Rightarrow X_\infty$ as $n \to \infty$. If $h : \mathbb{R}^d \to \mathbb{R}^l$ is continuous, then $h(X_n) \Rightarrow h(X_\infty)$ as $n \to \infty$.

Extended Continuous Mapping Principle: Suppose $(X_n : n \geq 1)$ is a sequence of $\mathbb{R}^d$-valued random variables for which $X_n \Rightarrow X_\infty$ as $n \to \infty$. Given $h : \mathbb{R}^d \to \mathbb{R}^l$, let $D_h = \{x \in \mathbb{R}^d : h \text{ is not continuous at } x\}$. If $P(X_\infty \in D_h) = 0$, then $h(X_n) \Rightarrow h(X_\infty)$ as $n \to \infty$.

For the current application, we let $\chi_n = n^{\frac{1}{2}} \frac{(\bar{Z}_n - EZ)}{\sigma}$, $\beta_n = \sigma/s_n$, and note that $(\chi_n, \beta_n) \Rightarrow (N(0,1), 1)$. Setting $h(x,y) = xy$, we therefore conclude (via the continuous mapping principle) that

$$h(\chi_n, \beta_n) = n^{1/2} \frac{(\bar{Z}_n - EZ)}{s_n} \Rightarrow N(0,1)$$

as $n \to \infty$.

- Step 4. Note that the event

$$\left\{ -z \leq n^{1/2} \frac{(\bar{Z}_n - EZ)}{s_n} \leq z \right\} = \left\{ EZ \in \left[\bar{Z}_n - z \frac{s_n}{\sqrt{n}}, \bar{Z}_n + z \frac{s_n}{\sqrt{n}} \right] \right\}$$

1.7 More on the Weak Law of Large Numbers

It is often of interest to generalize the weak law to correlated processes, e.g. Z_i = inventory level in period i. In a typical application, the Z_i's are dependent random variables. How can we establish that $\bar{Z}_n$ converges in such models?

Start with what is often a very reasonable assumption. Assume that $(Z_n : n \geq 0)$ is a *stationary sequence*, so that

$$(Z_0,\, Z_1, \dots) \stackrel{\mathcal{D}}{=} (Z_1,\, Z_2, \dots)$$

We hope to establish, in some generality, that

$$\bar{Z}_n \xrightarrow{p} EZ_0. \tag{1.7.1}$$

Remark 1.7.1 The limit (1.7.1) is not universally valid. Give an easy counterexample.

To prove (1.7.1), we need to show that for each $\epsilon > 0$,

$$P(|\bar{Z}_n - EZ_0| > \epsilon) \to 0$$

as $n \to \infty$. But Chebyshev's inequality implies that

$$P(|\bar{Z}_n - EZ_0| > \epsilon) \leq \frac{\mathrm{var}\, \bar{Z}_n}{\epsilon^2}.$$

Also,

$$\begin{aligned}
\mathrm{var}\, \bar{Z}_n &= \frac{1}{n^2} \mathrm{cov} \left(\sum_{i=1}^n Z_i,\, \sum_{j=1}^n Z_j \right) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathrm{cov}\,(Z_i,\, Z_j) \\
&= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathrm{cov}\,(Z_0,\, Z_{j-i}) \text{ (via stationarity)} \\
&= \frac{1}{n^2} \left(n \mathrm{var}\, Z_0 + 2 \sum_{i=1}^{n-1} (n-i) \mathrm{cov}\,(Z_0,\, Z_i) \right).
\end{aligned}$$

Hence, if $\mathrm{var}\, Z_0 < \infty$ and $\mathrm{cov}\,(Z_0,\, Z_i) \to 0$ as $n \to \infty$, it follows that $\mathrm{var}\, \bar{Z}_n \to 0$ as $n \to \infty$, proving that $\bar{Z}_n \xrightarrow{p} EZ_0$ as $n \to \infty$. So, all that is needed to establish (1.7.1) is a (possibly crude) bound on the covariance.

1.8 The Strong Law of Large Numbers

Another variant of the *law of large numbers* (LLN) is the so-called *strong law of large numbers* (SLLN).

Definition 1.8.1 We say that X_n *converges almost surely* to X_∞ (and write $X_n \xrightarrow{a.s.} X_\infty$ as $n \to \infty$) if $P(A) = 1$, where

$$A = \{\omega : X_n(\omega) \to X_\infty(\omega) \text{ as } n \to \infty\}$$

Remark 1.8.1 Almost sure convergence is also known as: convergence with probability one; convergence almost everywhere; convergence almost certainly.

Remark 1.8.2 Note that the truth of whether $\omega \in A$ depends on infinitely many of the X_i's. Consequently, $P(A)$ can not be computed via a calculation based on the "finite-dimensional" distributions of the sequence $(X_n : n \geq 1)$. Thus, computing $P(A)$ typically requires an ability to assign probabilities to "infinite-dimensional events". Measure-theoretic probability was created, in part, as a means of being able to rigorously assign probabilities to such events.

The following result is the basic statement of the SLLN.

Theorem 1.8.1 Suppose that $Z_1, Z_2, \ldots$ is an iid sequence of $\mathbb{R}^d$-valued random variables. Then

$$\bar{Z}_n \overset{a.s.}{\to} EZ_1$$

as $n \to \infty$.

Proving the strong law is reseaonably straightforward when the Z_i's are *bounded random variables* (i.e. there exists $c < \infty$ such that $P(\|Z_i\| \leq c) = 1$).

Problem 1.8.1 Suppose that $Z_1, Z_2, \ldots$ is an iid sequence of bounded scalar random variables.

1. Use Chebyshev's inequality to prove that for any integer $m \geq 1$,

$$E\sum_{n=1}^{\infty} \mathbf{1}\left(|\bar{Z}_{n^2} - EZ_1| > \frac{1}{m}\right) < \infty$$

2. Conclude from the previous part that

$$\bar{Z}_{n^2} \overset{a.s.}{\to} EZ_1$$

as $n \to \infty$.

3. Use the boundedness of the Z_i's to conclude that

$$\bar{Z}_n \overset{a.s.}{\to} EZ_1$$

as $n \to \infty$. (Hint: For $j \in \{n^2+1, \ldots, (n+1)^2\}$, find an upper bound on $|\bar{Z}_j - \bar{Z}_{n^2}|$.)

4. Use your scalar result to prove the SLLN for $\mathbb{R}^d$-valued random variables.

The above line of argument extends readily to certain dependent sequences.

Problem 1.8.2 Suppose that $Z_1, Z_2, \ldots$ is a stationary sequence of bounded random variables for which

$$\sum_{i=1}^{\infty} |\mathrm{cov}(Z_0, Z_i)| < \infty$$

1. Use Chebyshev's inequality to prove that for any integer $m \geq 1$,

$$E\sum_{n=1}^{\infty} \mathbf{1}\left(|\bar{Z}_{n^2} - EZ_1| > \frac{1}{m}\right) < \infty$$

2. Use the argument of problem 1.8.1 to conclude that

$$\bar{Z}_n \overset{a.s.}{\to} EZ_1$$

as $n \to \infty$.

One big advantage of almost sure convergence is that many results that are difficult to establish using weaker forms of convergence are often easy to argue in the almost sure setting. For example, it is typically not the case that if $\chi_n \Rightarrow \chi_\infty$ and $\beta_n \Rightarrow \beta_\infty$ that $\chi_n + \beta_n \Rightarrow \chi_\infty + \beta_\infty$. But this is easy to establish if the weak convergence is required by almost sure convergence. The basic idea is to reduce the conclusion to a real variables (deterministic) argument by arguing "path by path". In particular, the almost sure convergence of χ_n to χ_∞ and β_n to β_∞ implies that if

$$A_1 = \{\omega : \chi_n(\omega) \to \chi_\infty(\omega) \text{ as } n \to \infty\}$$

and

$$A_2 = \{\omega : \beta_n(\omega) \to \beta_\infty(\omega) \text{ as } n \to \infty\}$$

then $P(A_1) = P(A_2) = 1$. But real analysis makes clear that $A_1 \cap A_2 \subseteq B$, where

$$B = \{\omega : \chi_n(\omega) + \beta_n(\omega) \to \chi_\infty(\omega) + \beta_\infty(\omega) \text{ as } n \to \infty\}.$$

Consequently, $P(B) \geq P(A_1 \cap A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2) \geq P(A_1) + P(A_2) - 1 = 1$, so therefore

$$\chi_n + \beta_n \overset{a.s.}{\to} \chi_\infty + \beta_\infty$$

as $n \to \infty$. These "path by path" (or "*sample path*") arguments are used extensively in the analysis of stochastic systems.

As noted earlier, the strong law is a statement about an infinite-dimensional event. Assigning probabilities to such infinite-dimensional events is nontrivial.

Example 1.8.1 Consider the SLLN in which $Z_i \overset{\mathcal{D}}{=} Ber(1/2)$ for $i \geq 1$. A natural choice for Ω here is

$$\Omega = \{0,1\}^\infty,$$

so that $\omega \in \Omega$ takes the form $\omega = (z_1, z_2, \dots)$. Then, $Z_i(\omega) = z_i$ for $i \geq 1$, and $P(Z_1 = z_1, \dots, Z_n = z_n) = 2^{-n}$ when $z_i \in \{0,1\}$ for $i \geq 1$. A naive approach to computing $P(A)$ for

$$A = \{\omega : \bar{Z}_n(\omega) \to 1/2 \text{ as } n \to \infty\}$$

would be to calculate

$$\sum_{\omega \in A} P(Z_1 = z_1, Z_2 = z_2, \dots).$$

This approach fails for (at least) two reasons:

- $P(Z_1 = z_1, Z_2 = z_2, \dots) = 0$ for each sequence $(z_1, z_2, \dots)$;
- The number of ω's in A is uncountably infinite.

The assignment of probability to events like A requires the tool of measure theory. Measure theory guarantees that the probability assignment associated with P can be uniquely extended from "finite-dimensional assignments" (e.g. $P(Z_1 = z_1, \ldots, Z_n = z_n) = 2^{-n}$) to a class of infinite-dimensional events that includes A. We implicitly use this result all over the world of stochastic modeling (e.g. whenever we compute probabilities of infinite-dimensional events, having only unambiguously assigned probabilities to finite-dimensional characteristics of the model).

Remark 1.8.3 Measure theory does **not** tell us that we may uniquely assign probabilities to all infinite-dimensional event. Typically, there are so-called non-measurable events to which probabilities can not be assigned. But virtually all events that will naturally arise in applications are "measurable events" to which probabilities can be uniquely assigned. So, we will not worry about non-measurable events in this course.

Not surprisingly, a powerful tool like the SLLN can have some interesting application. Let U be a uniform random variable in $\Omega = [0, 1]$, and consider the (random) binary expansion for U:

$$U = \sum_{n=1}^{\infty} Z_n 2^{-n},$$

where $Z_i \in \{0, 1\}$. It is easy to see that the Z_i's are iid $Ber(\frac{1}{2})$ random variable. So, the SLLN guarantees that

$$\begin{aligned} & P(\text{long-run fraction of 0's, is the binary expansion of } u) \\ = \; & P(\text{long-run fraction of 1's, is the binary expansion of } u) \\ = \; & \frac{1}{2}. \end{aligned}$$

In other words, if $\Omega = [0, 1]$ and $Z_i(\omega) = i$'th binary coefficient in the binary expansion of $\omega \in [0, 1]$, then

$$P(\omega \in A) = 1$$

where $A = \{\omega : \bar{Z}_n(\omega) \to 1/2 \text{ as } n \to \infty\}$, provided that

$$P(\omega \leq x) = x \tag{1.8.1}$$

for $0 \leq x \leq 1$. ((1.8.1) is just asserting that P assigns probability uniformly.)

Supposed now that $W \in [0, 1]$ is a random variable having a positive pdf $f_W(\cdot)$ on $[0, 1]$. Then,

$$P(W \in A^c) = \int_{A^c} f_W(x) dx.$$

But we have already shown that

$$\int_{A^c} dx = 0,$$

so evidently $P(W \in A^c) = 0$, and hence $P(W \in A) = 1$. In other words, any "continuous random variable" on $[0, 1]$ also has the property that the long-run fraction of 0's in its binary representation is $\frac{1}{2}$.

Problem 1.8.3 Prove that if W is a random variable having a pdf $f_W(\cdot)$, then

$$P(\text{long-run fraction of i's in } W\text{'s decimal expansion}) = \frac{1}{10}$$

for $0 \leq i \leq 9$.

Remark 1.8.4 As a consequence of the above discussion, any random variable W constructed as follows:

$$W = \sum_{n=1}^{\infty} Z_n 2^{-n}$$

where the Z_i's are iid $Ber(p)$ (with $p = \frac{1}{2}$) cannot have a pdf. But such a W is clearly also not discrete. Such W's are random variables that have a continuous distribution function that is not absolutely continuous. These types of random variables are difficult to deal with computationally. Fortunately, from a modeling viewpoint, we can safely ignore them. (Why?)

1.9 The Ergodic Theorem

In many applications of interest, we will wish to develop strong laws in the dependent setting. The *ergodic theorem* (due to Birkhoff and Von Neumann) is one essential tool.

Theorem 1.9.1 Let $(Z_n : n \geq 0)$ be a stationary sequence for which $E|Z_0| < \infty$. Then, there exists W for which $EW = EZ_0$ and

$$\bar{Z}_n \overset{a.s.}{\to} W$$

as $n \to \infty$.

Remark 1.9.1 Note that the ergodic theorem guarantees that $\bar{Z}_n$ always has an a.s. limit, provided only that we require the Z_i's to be stationary and integrable. If W is deterministic, then clearly $W = EZ_0$ and we get the standard strong law conclusion. In view of this, we say that $(Z_n : n \geq 0)$ is *ergodic* if W is deterministic, and *non-ergodic* if W is random.

Problem 1.9.1 Give a (simple!) example where W is random.

Problem 1.9.2 Prove that if $(Z_n : n \geq 0)$ is a stationary sequence for which $EZ_0^2 < \infty$ and $\text{cov}(Z_0, Z_n) \to 0$ as $n \to \infty$, then $(Z_n : n \geq 0)$ is ergodic.

We now illustrate the application of this result in the setting of the *newsvendor model.* Consider a retailer that carries perishable inventory that can not be carried over to the next time period. (e.g. newspapers). What is the "optimal order quantity" x^* for the retailer?

Let $D_1, D_2, \ldots$ be the successive demands in periods $1, 2, \ldots$ and let x be the order quantity. If $c =$ cost/item, $r =$ revenue/item, and $s =$ sales value/item, then the profit earned in period i is

$$\text{Profit}_i(x) = r(D_i \wedge x) + s(x - (D_i \wedge x)) - cx.$$

Assuming that we have a stationary ergodic process, we find that

$$\frac{1}{n} \sum_{i=1}^{n} \text{Profit}_i(x) \overset{a.s.}{\to} E[r(D_i \wedge x) + s(x - (D_i \wedge x)) - cx]$$

as $n \to \infty$, so that

$$\sum_{i=1}^{n} \text{Profit}_i(x) \approx nE[r(D_i \wedge x) + s(x - (D_i \wedge x)) - cx]$$

for n large. Since the goal of the retailer is to maximize total profit, we should choose x to maximize the (deterministic!) expectation

$$E[r(D_i \wedge x) + s(x - (D_i \wedge x)) - cx].$$

Assuming that the demand D_0 is a continuous random variable with pdf $f_D(\cdot)$,

$$E[r(D_i \wedge x) + s(x - (D_i \wedge x)) - cx] = r\int_0^x y f_D(y)dy + rxP(D_1 \geq x) + s\int_0^x (x-y)f_D(y)dy - cx,$$

so that the optimizing x^* is a root of

$$P(D_1 \leq x^*) = \frac{r-c}{r-s}.$$

(Note that it is natural to assume that $0 < s < c < r$, so that $0 \leq (r-c)/(r-s) \leq 1$.) This optimal order quantity formula is one of the most important formulae in the area of operations management, and plays a role in many different applied settings (airline reservations, hotel rooms, electric power from renewables). Our argument establishes that the formulae continues to hold even in settings where the demands are correlated.

1.10 The Subadditive Ergodic Theorem

One way to motivate the ergodic theorem of Section 1.9 is to consider first a deterministic additive sequence $(s_{m,n} : 0 \leq m \leq n)$ satisfying:

i) $s_{0,0} = 0$

ii) $s_{0,m+n} = s_{0,m} + s_{m,m+n}$

iii) $s_{m,m+n} = s_{0,n}$.

It is easily seen from these postulates that $s_{0,n} = ns_{0,1}$, so that

$$\frac{1}{n}s_{0,n} \to s_{0,1} \tag{1.10.1}$$

as $n \to \infty$.

What might a stochastic analog of i) - iii) look like? We consider a stochastic sequence $(S_{m,n} : 0 \leq m \leq n)$ satisfying

i') $S_{0,0} = 0$

ii') $S_{0,m+n} = S_{0,m} + S_{m,m+n}$

Note that property ii') implies that

$$S_{0,n} = \sum_{i=1}^n \beta_i$$

where $\beta_i = S_{i-1,i}$. The stochastic analog of iii) is to require that

iii') $(\beta_i : i \geq 1)$ is a stationary sequence (i.e. $(\beta_{m+i} : i \geq 1) \stackrel{\mathcal{D}}{=} (\beta_i : i \geq 1)$ for $m \geq 1$).

Property iii') implies that

$$S_{m,\, m+n} \overset{\mathcal{D}}{=} S_{0,\, n},$$

which of course mirrors assumption iii) for the deterministic sequence. The ergodic theorem can be re-phrased as follows:

Proposition 1.10.1 Under conditions i') - iii') above,

$$\frac{1}{n} S_{0,\, n} \to Z \quad \text{a.s.}$$

as $n \to \infty$, where Z is a rv for which $\mathrm{E}Z = \mathrm{E}\beta_1$, provided that $\mathrm{E}|\beta_1| < \infty$.

A deterministic sequence $(s_{m,\, n} : 0 \le m \le n)$ is called *subadditive* if:

i) $s_{0,\, 0} = 0$

ii) $s_{0,\, m+n} \le s_{0,\, m} + s_{m,\, m+n}$

iii) $s_{m,\, m+n} = s_{0,\, n}$.

It is quite straightforward to prove that a subadditive sequence has the property that

$$\frac{1}{n} s_{0,\, n} \to \inf_{m \ge 1} \frac{s_{0,\, m}}{m} \tag{1.10.2}$$

as $n \to \infty$.

Exercise 1.10.1 Prove (1.10.2).

The Subadditive Ergodic Theorem is a stochastic analog to (1.10.2). The generalization bears great similarity to the Ergodic Theorem. Specially, let $(S_{m,\, n} : 0 \le m \le n)$ be a family of rv's satisfying:

i') $S_{0,\, 0} = 0$

ii') $S_{0,\, m+n} \le S_{0,\, m} + S_{m,\, m+n}$

iii') $(S_{m,\, m+k} : k \ge 0) \overset{\mathcal{D}}{=} (S_{m+l,\, m+l+k} : k \ge 0)$ for $l \ge 1$

iv') $(S_{mk,\, (m+1)k} : m \ge 0)$ is stationary for $k \ge 1$.

Observe that iii') - iv') is a natural stochastic analog to property iii) for a deterministic subadditive sequence.

Theorem 1.10.1 (The Subadditive Ergodic Theorem) Suppose that $(S_{m,\, n} : 0 \le m \le n)$ satisfies i')-iv') above. If $\mathrm{E}S_{0,\, 1}^{+} < \infty$, then

$$\frac{1}{n} S_{0,\, n} \to Z \quad \text{a.s.}$$

as $n \to \infty$, where $-\infty \le Z < \infty$ a.s. and

$$\mathrm{E}Z = \inf_{m \ge 1} \frac{1}{m} \mathrm{E}S_{0,\, m}. \tag{1.10.3}$$

As in the ergodic theorem, ergodicity guarantees that Z is deterministic. Specially, if $(S_{mk,\, (m+1)k} : m \ge 0)$ is ergodic for each $k \ge 1$, then Z is necessarily deterministic.