Chapter 3

RANDOM ELEMENTS

1 Introduction

This chapter consists of two parts: Sections 2–5 and Sections 6–10. The first part introduces general coupling tools, and the second part generalizes some of the results from Chapters 1 and 2.

After a measure-theoretic review of terminology in Section 2, Section 3 explains what is meant by extending the underlying probability space and collects some extension techniques. Sections 4 and 5 are devoted to particularly important extension techniques: conditioning, transfer, and splitting. These sections may seem rather technical, but the extension methods are quite probabilistic in nature, and are used frequently throughout the book.

In Section 6, transfer and splitting are used to construct a successful coupling of random walks with step-lengths that are spread out (continuous step-lengths are a special case of spread out). In Section 7, splitting is used to construct a maximal coupling of an arbitrary collection of random elements. In Section 8, we consider the special case of two random elements and formulate the maximal coupling result in terms of total variation. In Section 9, splitting is used to turn lim inf convergence of distributions to a distribution into a pointwise convergence where the random elements actually hit the limit and stay there. In Section 10, we use transfer (and Theorem 6.1 in Chapter 1) to turn convergence in distribution into pointwise convergence for random elements in a separable metric space (Dudley's extension of the Skorohod coupling), and then re-prove this result in the case when the space is also complete (the Skorohod coupling) after generalizing of the quantile coupling.

2 Back to Basics – Definition of Coupling

A *random element* in a measurable space $(E, \mathcal{E})$ *defined on* a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is a measurable mapping Y from $(\Omega, \mathcal{F}, \mathbf{P})$ to $(E, \mathcal{E})$, that is,

$$\{Y \in A\} \in \mathcal{F}, \quad A \in \mathcal{E},$$

where

$$\{Y \in A\} := \{\omega \in \Omega : Y(\omega) \in A\} =: Y^{-1}A.$$

We also say that Y is *supported by* $(\Omega, \mathcal{F}, \mathbf{P})$ and that Y is an $\mathcal{F}/\mathcal{E}$ measurable mapping from Ω to E. Note that if we replace $\mathbf{P}$ by another probability measure $\mathbf{Q}$, then Y is the same measurable mapping but a different random element. Also, if we replace $\mathcal{F}$ by a larger σ-algebra and/or $\mathcal{E}$ by a smaller, then Y is the same mapping but not the same measurable mapping. If we replace $\mathcal{F}$ by a smaller σ-algebra and/or $\mathcal{E}$ by a larger, then Y need not even be measurable.

A *random variable* X is a random element in $(\mathbb{R}, \mathcal{B})$, where $\mathbb{R}$ denotes the set of real numbers (the line) and $\mathcal{B}$ its *Borel* subsets, that is, $\mathcal{B}$ is the smallest σ-algebra on $\mathbb{R}$ containing the open sets (the σ-algebra *generated* or *induced* by the open sets),

$$\mathcal{B} = \mathcal{B}(\mathbb{R}) = \sigma\{A \subseteq \mathbb{R} : A \text{ open}\}.$$

An *extended* random variable X is a random element in

$$([-\infty, \infty], \mathcal{B}([-\infty, \infty])).$$

When the line is regarded as time, we often call an extended random variable T a *random time.* If T cannot take the values $-\infty$ and ∞, then T is a *finite* random time.

The *distribution* of a random element Y [under $\mathbf{P}$] is the probability measure on $(E, \mathcal{E})$ *induced* by Y, namely $\mathbf{P}Y^{-1}$. Since

$$\mathbf{P}(Y \in A) = \mathbf{P}Y^{-1}A, \quad A \in \mathcal{E},$$

a more probabilistic notation for the distribution of Y is $\mathbf{P}(Y \in \cdot)$.

A random element Y is *canonical* if Y is the identity mapping, that is, if

$$(\Omega, \mathcal{F}) = (E, \mathcal{E}) \quad \text{and} \quad Y(\omega) = \omega, \quad \omega \in \Omega.$$

Then $\mathbf{P}(Y \in \cdot) = \mathbf{P}$.

A random element $\hat{Y}$ in $(E, \mathcal{E})$ defined on a probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbf{P}})$ is a *copy* or *representation* of Y if

$$\hat{\mathbf{P}}(\hat{Y} \in \cdot) = \mathbf{P}(Y \in \cdot); \quad \text{this is denoted by} \quad \hat{Y} \stackrel{D}{=} Y.$$

A random element Y has always a *canonical representation*, the canonical random element on $(E, \mathcal{E}, \mathbf{P}(Y \in \cdot))$.

2.1 Coupling Random Elements – Definition

For each i in an index set $\mathbb{I}$ let Y_i be a random element in a measurable space $(E_i, \mathcal{E}_i)$ defined on a probability space $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$. A family of random elements $(\hat{Y}_i : i \in \mathbb{I})$ defined on a common probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbf{P}})$ is a *coupling* of Y_i, $i \in \mathbb{I}$, if

$$\hat{Y}_i \stackrel{D}{=} Y_i$$

for each $i \in \mathbb{I}$.

Note that the Y_i need not be defined on a common probability space, in other words need not have a joint distribution. Thus 'coupling' can be seen to refer to the fact that the copies $\hat{Y}_i$ are defined on a common probability space, have a joint distribution, live together. Writing $(\hat{Y}_i : i \in \mathbb{I})$ in parentheses indicates that this is the case.

For any collection of random elements Y_i, $i \in \mathbb{I}$, there is always at least one coupling, the *independence coupling*, consisting of independent copies of the Y_i. This follows from the product measure theorem (Fact 3.1 below).

2.2 Coupling Probability Measures – Rephrasing the Definition

In terms of distributions the definition of coupling can be rephrased as follows. For each i in an index set $\mathbb{I}$ let P_i be a probability measure on a measurable space $(E_i, \mathcal{E}_i)$. Define the *product space*

$$\bigotimes_{i \in \mathbb{I}} (E_i, \mathcal{E}_i) := \Big(\prod_{i \in \mathbb{I}} E_i, \bigotimes_{i \in \mathbb{I}} \mathcal{E}_i \Big)$$

where $\prod_{i \in \mathbb{I}} E_i$ is the *Cartesian product* of the E_i,

$$\prod_{i \in \mathbb{I}} E_i := \{ y = (y_i : i \in \mathbb{I}) : y_i \in E_i, i \in \mathbb{I} \},$$

and $\bigotimes_{i \in \mathbb{I}} \mathcal{E}_i$ is the *product σ-algebra*, that is, the smallest σ-algebra on $\prod_{i \in \mathbb{I}} E_i$ making the ith *projection mapping* taking y in $\prod_{i \in \mathbb{I}} E_i$ to y_i in E_i measurable for all $i \in \mathbb{I}$ (the σ-algebra *generated* or *induced* by the projection mappings):

$$\bigotimes_{i \in \mathbb{I}} \mathcal{E}_i := \sigma\{\{y : y_i \in A\} : i \in \mathbb{I} \text{ and } A \in \mathcal{E}_i\}.$$

A probability measure P on $\bigotimes_{i \in \mathbb{I}} (E_i, \mathcal{E}_i)$ is a *coupling* of P_i, $i \in \mathbb{I}$, if P_i is the ith *marginal* of P, that is, if P_i is induced by the ith projection mapping:

$$P(\{y : y_i \in A\}) = P_i(A), \quad A \in \mathcal{E}_i, \quad i \in \mathbb{I}.$$

2.3 The Relation Between the two Formulations

The latter definition of coupling can be seen as a *canonical version* of the former by the following identification: let $(E_i, \mathcal{E}_i, P_i)$ be the probability spaces supporting the canonical copies of the individual random elements Y_i, that is,

$$P_i = \mathbf{P}_i(Y_i \in \cdot), \quad i \in \mathbb{I},$$

and let $(\prod_{i\in\mathbb{I}} E_i, \bigotimes_{i\in\mathbb{I}} \mathcal{E}_i, P)$ be the probability space supporting the canonical copy of the coupling $(\hat{Y}_i : i \in \mathbb{I})$, that is,

$$P = \hat{\mathbf{P}}((\hat{Y}_i : i \in \mathbb{I}) \in \cdot).$$

Note that here we treat the expression $(\hat{Y}_i : i \in \mathbb{I})$ not as a collection of individual random elements in $(E_i, \mathcal{E}_i)$, $i \in \mathbb{I}$, but as a single random element in $\bigotimes_{i\in\mathbb{I}}(E_i, \mathcal{E}_i)$ defined by

$$(\hat{Y}_i : i \in \mathbb{I})(\hat{\omega}) := (\hat{Y}_i(\hat{\omega}) : i \in \mathbb{I}), \quad \hat{\omega} \in \hat{\Omega}.$$

The distribution of this random element, $\hat{\mathbf{P}}((\hat{Y}_i : i \in \mathbb{I}) \in \cdot)$, is the *joint distribution* of the $\hat{Y}_i$, $i \in \mathbb{I}$.

3 Extension Techniques

Finding random elements with particular properties (having a particular joint distribution with those already introduced) can be an essential task in constructing couplings. These new random elements are often brought into existence *by extension*, by extending the underlying probability space. In this section, and Sections 4 and 5, we give several ways of doing this. Let us start by making precise what we mean by extension.

3.1 Extending the Underlying Probability Space – Definition

A probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}})$ is an *extension* of another probability space $(\Omega, \mathcal{F}, \mathbf{P})$ if $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}})$ supports a random element ξ in $(\Omega, \mathcal{F})$ having $\mathbf{P}$ as distribution. If Y is a random element in $(E, \mathcal{E})$ defined on $(\Omega, \mathcal{F}, \mathbf{P})$, then the random element $\bar{Y}$ defined on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}})$ by

$$\bar{Y}(\bar{\omega}) = Y(\xi(\bar{\omega})), \quad \bar{\omega} \in \bar{\Omega}, \qquad \text{(see Figure 3.1)}$$

is a copy of Y, since for $A \in \mathcal{E}$,

$$\begin{aligned}\bar{\mathbf{P}}(\bar{Y} \in A) &= \bar{\mathbf{P}}(\xi \in Y^{-1}A) \\ &= \mathbf{P}(Y^{-1}A) = \mathbf{P}(Y \in A).\end{aligned}$$

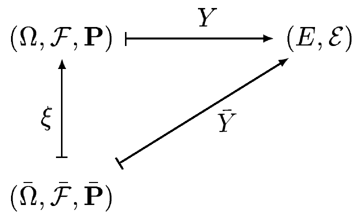

FIGURE 3.1. The *original* random element $\bar{Y}$ *induced* by Y.

Say that $\bar{Y}$ is *induced* by Y and call $\bar{Y}$ an *orginal* random element. Thus, in particular, ξ is the original random element induced by the canonical random element on $(\Omega, \mathcal{F}, \mathbf{P})$. In addition to the original random elements (which we shall think of as 'old') the probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}})$ may support 'new' random elements not induced by random elements already supported by $(\Omega, \mathcal{F}, \mathbf{P})$. These 'new' random elements we shall call *external*.

Convention of the common probability space. When there is no risk of confusion, we often extend the underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$ without changing its name: after extending $(\Omega, \mathcal{F}, \mathbf{P})$ to obtain $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}})$ we rename the extension $(\Omega, \mathcal{F}, \mathbf{P})$, and the induced $\bar{Y}$ we rename Y. This identification of $\bar{Y}$ and Y explains the term 'original' for $\bar{Y}$. This procedure enables us to assume, when convenient, that all the random elements to be considered in a certain context are defined on a common probability space, which we then denote by $(\Omega, \mathcal{F}, \mathbf{P})$. Call this the *convention of the common probability space*. Typically, in probability theory, it is not the actual underlying probability space that matters but the (joint) distributions of the random elements under consideration.

It is, however, crucial that new random elements be introduced in a consistent manner and, as the example in Section 10 of Chapter 1 shows, we must be careful here. For the rest of this section, and in the next two, we give several safe ways of introducing new random elements. But first we consider at some length an extension that does not yield any new random elements.

3.2 Reduction Extension – Deleting a Null Event

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. An element $\omega \in \Omega$ is an *outcome* and a set $A \in \mathcal{F}$ is an *event*. If $\mathbf{P}(A) = 0$, then A is a *null* event. If $\mathbf{P}(A) = 1$, then A is an *almost sure* (a.s.) event. Any statement that holds on an a.s. event (that is, for all outcomes in the event) is an a.s. statement.

It is common practice to remove a null event (or a set contained in a null event, an *outer null set*) from the underlying probability space, thereby