The Annals of Applied Probability
1995, Vol. 23, No. 3, 1456–1460

BOOK REVIEW

TORGNY LINDVALL, *Lectures on the Coupling Method.* Wiley, New York, 1992.

REVIEW BY NEIL O'CONNELL

Dublin Institute for Advanced Studies

This really is a delightful book. For me, not being an expert, it was as though the author had set out to "charm me with his subject," and like a good teacher, he succeeded. The exposition is lively, unassuming, highly motivating, and conscientious, similar in spirit to the books of Williams [5, 6]. The subject is taught by example. In the words of the author:

> To know a method is to have learned how it works. What we have ahead of us is essentially a collection of applications of a few basic ideas consisting largely of topics of wide common interest with an attempt to maximize diversity.

It is intended to serve graduate courses and seminars in departments of mathematics, statistics and operations research. Basic familiarity with measure-theoretic probability is assumed.

Written primarily as a textbook, it is also the first definitive reference on the coupling method and its many uses. From a practical point of view it is well signposted and, well, "mixing" in the sense that it can be opened and understood to at least some extent at any point in the text. For a subject with such far-reaching applications, the author has done well to include so many. It seems to me that what is not included is at least hinted at and the reader is well informed on where to look for more details.

We will now present a brief introduction to the coupling method, just to pique your interest. This will be followed by an outline of the text.

The coupling method. The coupling method provides an ingenious way of comparing probability measures. The idea is to construct, if possible, random elements on a common probability space in such a way that the comparison may be carried out in terms of the random elements. Roughly speaking, applications fall into three categories:

1. Establishing ergodicity and obtaining rates of convergence to stationarity for Markov processes.
2. Justifying weak approximations.
3. Establishing inequalities.

The first two are achieved via estimates of total variation distances, based on the *coupling inequality.* (The coupling inequality for Markov chains is described below: we refer the reader to the text for a general version.)

Received October 1994

To illustrate the method, we have selected three typical applications, one from each of the above categories.

The coupling inequality for Markov chains. Consider an aperiodic, irreducible, positive recurrent Markov chain $X = (X_n)$ on the nonnegative integers with transition matrix $P = (p_{ij})$ and (unique) stationary distribution $\pi = (\pi_j)$. It is well known that, regardless of the initial distribution λ, X_n approaches stationarity as $n \to \infty$. In other words, $\lambda P^n \to \pi$. This is a classical result. To prove it by coupling, we introduce a stationary version of X: Let X' be a Markov chain independent of X, with transition matrix P and initial distribution π. We now couple X and X' at the stopping time

$$T = \min\{k\colon\ X_k = X'_k\}$$

by defining a new process

$$X''_n = \begin{cases} X_n, & n < T, \\ X'_n, & n \geq T. \end{cases}$$

By the strong Markov property, X and X'' are equally distributed, and so

$$\begin{aligned} |P(X_n = j) - \pi_j| &= |P(X''_n = j) - P(X'_n = j)| \\ &= |P(X''_n = j,\ T \leq n) + P(X''_n = j,\ T > n) \\ &\quad - P(X'_n = j,\ T \leq n) - P(X'_n = j,\ T > n)| \\ &\leq P(X''_n = j,\ T > n) + P(X'_n = j,\ T > n). \end{aligned}$$

Summing over j yields

$$\|\lambda P^n - \pi\| \equiv \sum_j |P(X_n = j) - \pi_j| \leq 2P(T > n). \tag{1}$$

This is a special case of the coupling inequality. It follows from (1) that if the coupling is successful, we have uniform convergence to stationarity. (The coupling is "successful" if T is almost surely finite.) The coupling inequality also provides a route for obtaining rates of convergence.

The Stein–Chen method. In order to apply the coupling method for the comparison of two probability measures, one must first find a suitable coupling: this is often the hard part. The challenge of cooking up a clever coupling is what makes the method so attractive, and has in many applications produced some beautiful mathematics: the Stein–Chen method for Poisson approximation is such an example.

Let $(Y_k,\ k = 1,\dots,n)$ be a sequence of Bernoulli random variables, not necessarily independent, with $p_k = P(Y_k = 1)$. Set $W = \sum_k Y_k$ and $\lambda = \sum_k p_k$. The idea is that if the Y_k's are only weakly dependent and the p_k's are small, the distribution of W should be approximately Poisson with mean λ. To formulate this, for each k let U_k and V_k be random variables on the same probability space, satisfying

$$U_k =_d W, \qquad 1 + V_k =_d P(W \in \cdot \mid Y_k = 1), \tag{2}$$

with the convention that $V_k = 0$ if $p_k = 0$. Assuming that the p_k's are small, the notion of weak dependence can now be expressed as "for each k, $U_k = V_k$ with high probability." Using the above coupling, one can show that

$$\|P(W \in \cdot) - p_\lambda\| \le 2(1 \wedge \lambda^{-1}) \sum_k p_k E|U_k - V_k|, \tag{3}$$

where p_λ denotes a Poisson distribution with mean λ. The idea of the proof is as follows. First note that if Z is Poisson with mean λ, then

$$E[\lambda h(Z+1) - Zh(Z)] = 0 \tag{4}$$

for any bounded function h. This suggests that if $E[\lambda h(W) - Wh(W)]$ is small, then the distribution of W is approximately Poisson with mean λ. So for each set A we find a function g for which

$$P(W \in A) - p_\lambda(A) = E[\lambda g(W+1) - Wg(W)], \tag{5}$$

namely, the solution to the equation

$$\lambda g(k+1) - kg(k) = I_A(i) - p_\lambda(A), \qquad k \ge 0, \tag{6}$$

with $g(0) = 0$. It can be shown that for any $j, k \ge 0$,

$$|g(j) - g(k)| \le (1 \wedge \lambda^{-1})|j - k|, \tag{7}$$

and so we have

$$\begin{aligned} |P(W \in A) - p_\lambda(A)| &= |E[\lambda g(W+1) - Wg(W)]| \\ &= \left|\sum_k p_k E[g(U_k+1) - g(V_k+1)]\right| \\ &\le (1 \wedge \lambda^{-1}) \sum_k p_k E|U_k - V_k|. \end{aligned}$$

The result now follows from the fact that

$$\|P(W \in \cdot) - p_\lambda\| = 2 \sup_A |P(W \in A) - p_\lambda(A)|. \tag{8}$$

The estimate (3) is in fact a generalization of the celebrated Le Cam theorem, which states that for independent Y_k's,

$$\|P(W \in \cdot) - p_\lambda\| \le 2 \sum_k p_k^2/\lambda \le 2 \max_k p_k.$$

To see this, note that in this case $E|U_k - V_k| = p_k$ and $1 \wedge \lambda^{-1} \le \lambda^{-1}$.

For variations of this result and other applications of the Stein–Chen method, we refer the reader to the text (and references therein) and the book of Barbour, Holst and Jansson [1].

Domination. The idea of using the coupling method to establish inequalities has been applied in a variety of situations. We have taken a simple example from the introduction of the text, where the idea can be used to prove something which is intuitively obvious, but otherwise hard to prove. Consider two independent, recurrent birth and death processes X and X' with the same transition rates and suppose $X_0 < X'_0$. Just as in the first example, we couple X and X' at

$$T = \inf\{t\colon\ X_t = X'_t\}$$

and define a new process X'' by

$$X''_t = \begin{cases} X_t, & t < T, \\ X'_t, & t \geq T. \end{cases}$$

Now X and X'' have the same distribution and, because we have skip-free paths, $X''_t \leq X'_t$ for all $t \geq 0$. It follows that for any nondecreasing function g, $Eg(X_t) \leq Eg(X'_t)$ for all $t \geq 0$. As the author puts it, "What alternative proof of that do you support?"

The author attributes the idea of coupling to Doeblin, who published a paper [2] in 1938 where the basic ergodic theorem for discrete state space Markov chains was proved using the coupling method. However, the method did not become established until it was found to be an indispensable tool in the study of interacting particle systems, which really took off in the 1970's. The author regards the papers of Pitman [4] and Griffeath [3] as "crucial for the awareness of the method."

Outline of *Lectures on the Coupling Method.* The body of the text is divided into six chapters:

I. Preliminaries
II. Discrete theory
III. Continuous theory
IV. Inequalities
V. Intensity-governed processes
VI. Diffusions

Chapter I is foundational. The term "coupling" is defined and the coupling inequality established in its most general form. The idea of a "maximal" coupling is introduced: Roughly speaking this is a coupling that achieves equality in the coupling inequality.

Chapter II starts with an account of coupling of discrete renewal processes, including a proof of the discrete renewal theorem and some rate results. There is not a renewal equation in sight, an omission which the author announces to be intentional. Renewal theory is a major theme throughout the book. Consequences of the coupling inequality for Markov chains are discussed, and the remainder of the chapter is devoted to examples in card-shuffling, random walks and Poisson approximation.

In Chapter III, the discrete renewal theory results of Chapter II are generalized to continuous time and to regenerative processes, and coupling is applied to Markov chains on a general state space. The rest of the chapter is concerned with questions regarding maximal coupling and ergodicity for Markov processes.

Chapter IV has Strassen's theorem as its basic result, which, roughly speaking, states that for a pair of stochastically ordered probability measures (on a partially ordered space E) there exists an associated coupling of almost surely ordered random elements (of E). The chapter includes a "gallery" of examples where the coupling method can be used to establish inequalities, ranging from percolation to Bernstein polynomials.

Chapter V demonstrates the use of coupling in the study of Markov processes which are defined in terms of transition intensities, such as general birth and death processes, interacting particle systems and renewal processes. There is also a section on embedding in Poisson processes, with applications to urn models and renewal processes.

The last chapter is devoted to the coupling of diffusions. For one-dimensional diffusions, a straightforward coupling (analogous to the one described in the Markov chain example above) will, due to path continuity, be successful. In higher dimensions the problem is not so easy, and one is faced with the challenge of finding an alternative coupling which will be successful. The author presents some examples where certain coupling can be guaranteed, including multidimensional Brownian motion and diffusions with radial drift and constant dispersion (variance).

REFERENCES

[1] BARBOUR, A., HOLST, L. and JANSSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press.

[2] DOEBLIN, W. (1938). Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkanique* **2** 77–105.

[3] GRIFFEATH, D. (1978). Coupling methods for Markov processes. In *Studies in Probability and Ergodic Theory. Advances in Mathematics Supplementary Studies* **2** (G.-C. Rota, ed.). Academic Press, New York.

[4] PITMAN, J. W. (1974). Uniform rates of convergence for Markov chain transition probabilities. *Z. Wahrsch. Verw. Gebiete* **29** 193–227.

[5] WILLIAMS, D. (1979). *Diffusions, Markov Processes and Martingales 1: Foundations*. Wiley, New York.

[6] WILLIAMS, D. (1991). *Probability with Martingales*. Cambridge Univ. Press.

DUBLIN INSTITUTE FOR ADVANCED STUDIES
10, BURLINGTON ROAD
DUBLIN 4
IRELAND