

Coupling I: concepts and examples

Wilfrid S. Kendall

Department of Statistics, University of Warwick

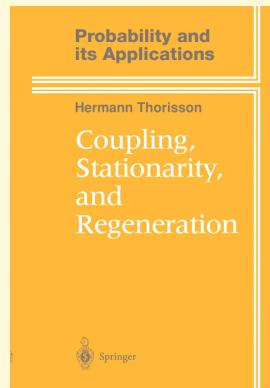
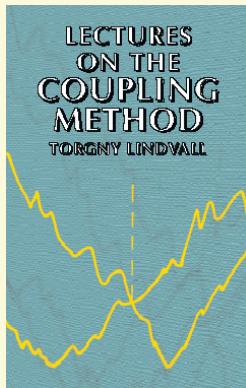
26 July 2003

Abstract

“Coupling” is a many-valued term in mathematical science! In a probabilist’s vocabulary it means: finding out about a random system X by constructing a second random system Y on the same probability space (maybe augmented by a seasoning of extra randomness). Careful construction, choosing the right system Y , designing the right kind of dependence between X and Y , can lead to clear intuitive explanations of important facts about X .

1 Introduction

- History: Doebelin (1938), see also Lindvall (1991).
- Literature: Breiman (1992), Lindvall (2002), Thorisson (2000), Aldous and Fill (200x);



- <http://www.warwick.ac.uk/statsdept/staff/WSK/talks/durham-lms1.pdf>
<http://research.microsoft.com/~dbwilson/exact/>

[Home Page](#)[Title Page](#)[Page 3 of 27](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Probability theory has a right and a left hand – Breiman (1992, Preface).

Coupling: more a probabilistic sub-culture than an identifiable theory.

A proof using coupling is rather like a well-told joke: if it has to be explained then it loses much of its force.

Coupling arguments are like counting arguments — but without natural numbers.



2 Monotonicity

2.1 Rabbits

Coupling for competing myxomatosis epidemics in Australian rabbits: Kendall and Saunders (1983).

$$\begin{aligned} s' &= -\alpha_1\beta_1si_1 - \alpha_2\beta_2si_2 \\ i_1' &= \alpha_1\beta_1si_1 - \beta_1i_1, & r_1' &= \beta_1i_1 \\ i_2' &= \alpha_2\beta_2si_2 - \beta_2i_2, & r_2' &= \beta_2i_2 \end{aligned}$$

Suppose $\alpha_1 > \alpha_2$. Are $r_1(\infty)$, $r_2(\infty)$ appropriately monotonic in i_1 , i_2 ?



[Home Page](#)[Title Page](#)

Page 5 of 27

[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Stochastic model. List potential infections from each individual as times-from-infection (*nb*: different infection rates for type-1 and type-2).

Converting type-1 initial infective to susceptible or to type-2 infective “clearly” delays progress of type-1 infections: hence desired monotonicity for stochastic model.

Deterministic model is limiting case, so inherits monotonicity.



Just one out of many applications to epidemic theory: another example is [Ball and Donnelly \(1995\)](#). For coupling in spatial epidemics, try [Mollison \(1977\)](#), [Häggström and Pemantle \(1998\)](#), [Häggström and Pemantle \(2000\)](#).

2.2 Igloos

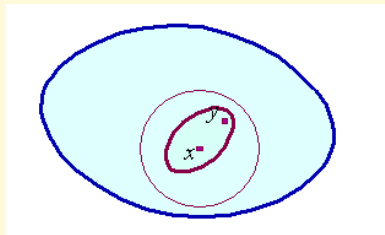


Are larger igloos always colder? Chavel (1986), Kendall (1989), Bass and Burdzy (1993). (See also Carmona and Zheng (1994), Hsu (1994).)

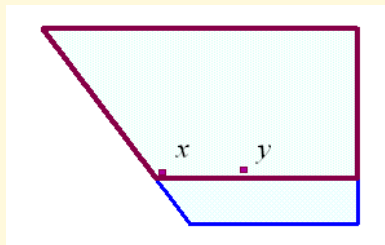
Light a match at $t = 0$ at location x . Compare temperatures $p_t(x, y)$, $q_t(x, y)$ at y in well-insulated igloos P and Q , where $P \subseteq Q$. Is it always colder in the larger igloo Q ?

- Yes for large times (convergence to uniformity);
- Not for all times for general igloos;
- Perhaps yes if the igloo is convex?

Move from Neumann heat kernel to reflecting Brownian motion.
Monotonicity does hold if there is a separating circle (WSK):



Monotonicity does not hold in general: obtuse angles cause trouble (B-B):



3 Representation

3.1 Split chains and small sets

Let X be a Markov chain on a non-discrete state space S , transition kernel $p(x, dy)$. Suppose a set C exists such that for some probability measure ν and some $\alpha > 0$

$$p(x, dy) \geq \mathbb{I}[C](x) \times \alpha \nu(dy). \quad (1)$$

Then X can be represented using new Markov chain on $S \cup \{\mathbf{c}\}$, for \mathbf{c} a regenerative pseudo-state.

(Idea due to Doeblin; for details see [Nummelin \(1978\)](#), also [Athreya and Ney \(1978\)](#).)

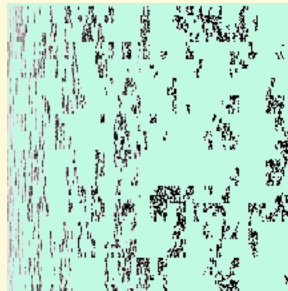
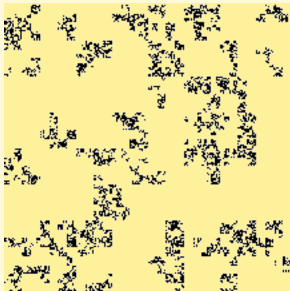
Higher-order small sets ($p(x, dy) \rightarrow p^{(k)}(x, dy)$) systematically reduce general state space theory to discrete. See [Meyn and Tweedie \(1993\)](#) also [Roberts and Rosenthal \(2001\)](#).

[Home Page](#)[Title Page](#)

Page 9 of 27

[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Small sets of order 1 need not exist: but will if (a) the kernel $p(x, dy)$ has a density and (b) chain is sub-sampled at *even* times.



Theorem 1 (*Kendall and Montana 2002*) *If the Markov chain has a measurable transition density $p(x, y)$ then the two-step density $p^{(2)}(x, y)$ can be expressed (non-uniquely) as a non-negative countable sum*

$$p^{(2)}(x, y) = \sum_i f_i(x)g_i(y).$$

Proof: Key Lemma, variation on Egoroff's Theorem:

Let $p(x, y)$ be an integrable function on $[0, 1]^2$. Then we can find subsets $A_\varepsilon \subset [0, 1]$, increasing as ε decreases, such that

- (a) for any fixed A_ε the " L^1 -valued function" p_x is uniformly continuous on A_ε : for any $\eta > 0$ we can find $\delta > 0$ such that $\int_0^1 |p_x(z) - p_{x'}(z)| dz < \eta$ for $|x - x'| < \delta$ and $x, x' \in A_\varepsilon$.
- (b) every point x in A_ε is of full relative density: as $u, v \rightarrow 0$ so $\text{Leb}([x - u, x + v] \cap A_\varepsilon) / (u + v) \rightarrow 1$.

□

[Home Page](#)[Title Page](#)[Page 11 of 27](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

3.2 FK representation and FKG

(See Sokal's talk)

4 Approximation

4.1 Skorokhod representation for weak convergence

Represent a random variable X with distribution function

$$F(x) = \mathbb{P}[X \leq x]$$

using the *inverse probability transform*

$$X = F^{-1}(U),$$

for U a Uniform $[0, 1]$ random variable. Do this with a single U for an entire weakly convergent sequence of random variables, to convert weak convergence to almost sure convergence.

(Extends to random variables with values in Polish spaces)

4.2 Central Limit Theorem and embedding in Brownian path

A zero-mean random variable X of finite variance can be represented as $X = B(T)$ for T a stopping time of finite mean. This allows us to deduce the (functional!) CLT from the Strong Law of Large Numbers and Brownian scaling.

$$X_n^{(N)} = B(\sigma_1^{(N)} + \dots + \sigma_n^{(N)}) - B(\sigma_1^{(N)} + \dots + \sigma_{n-1}^{(N)})$$

$$\frac{\sigma_1^{(N)} + \dots + \sigma_n^{(N)}}{N \mathbb{E}[\sigma]} \rightarrow 1 \quad \text{almost surely}$$

$$B\left(\frac{\sigma_1^{(N)} + \dots + \sigma_n^{(N)}}{N \mathbb{E}[\sigma]}\right) \rightarrow B(1) \quad \text{almost surely}$$

$$\text{LHS} \sim \frac{1}{\sqrt{N \mathbb{E}[\sigma]}} B(\sigma_1^{(N)} + \dots + \sigma_n^{(N)}) = \frac{1}{\sqrt{N \mathbb{E}[\sigma]}} \sum_1^N X_n^{(N)}$$

4.3 Stein-Chen method for Poisson approximation

Consider W a \mathbb{Z}^+ -valued random variable thought to be approximated by a $\text{Poisson}(\lambda)$ random variable \widetilde{W} .

Fix $A \subset \mathbb{Z}^+$. For $n \geq 0$, set

$$\lambda g(n+1) = ng(n) + \mathbb{I}[n \in A] - \mathbb{P}[\widetilde{W} \in A]. \quad (2)$$

One solution:

$$\frac{g(n+1) \times \mathbb{P}[\widetilde{W} = n+1] - \mathbb{P}[\widetilde{W} \in A, \widetilde{W} < n+1] + \mathbb{P}[\widetilde{W} \in A] \mathbb{P}[\widetilde{W} < n+1]}{n+1} \quad (3)$$

Hence (Barbour, Holst, and Janson 1992)

$$\sup_n |g(n)| \leq \min \left\{ 1, \frac{1}{\sqrt{\lambda}} \right\} \quad (4)$$

$$\sup_n |g(n+1) - g(n)| \leq \frac{1 - e^{-\lambda}}{\lambda} \leq \min \left\{ 1, \frac{1}{\lambda} \right\} \quad (5)$$

Also for any \mathbb{Z}^+ -valued random variable W ,

$$\mathbb{E}[\lambda g(W+1) - Wg(W)] = \mathbb{P}[W \in A] - \mathbb{P}[\widetilde{W} \in A] \quad (6)$$

Suppose $W = \sum I_i$ for dependent binary I_i .

$$\mathbb{E}[I_i g(W_i + 1)] = p_i \mathbb{E}[g(W)|I_i = 1]$$

where $W_i = \sum_{j \neq i} I_j$ and $p_i = \mathbb{E}[I_i]$ so $\sum_i p_i = \lambda$.

$$\mathbb{P}[W \in A] - \mathbb{P}[\widetilde{W} \in A] = \sum p_i (\mathbb{E}[g(W + 1)] - \mathbb{E}[g(W)|I_i = 1]).$$

Search for *coupled* U_i, V_i such that U_i has distribution of W , $V_i + 1$ has distribution of W given $I_i = 1$.

$$\left| \mathbb{P}[W \in A] - \mathbb{P}[\widetilde{W} \in A] \right| \leq \sup_n |g(n+1) - g(n)| \sum p_i \mathbb{E}[|U_i - V_i|]. \quad (7)$$

Even better if $U_i \leq V_i$ (say), giving collapse of sum:

$$\sum p_i \mathbb{E}[|U_i - V_i|] = \mathbf{Var}[W] - \lambda. \quad (8)$$

Easy example: case of independent $I_i \dots$

Based on example in Barbour et al. (1992, §2.3).

Poisson point process on a sphere, total intensity α . Consider short distances (Silverman and Brown 1978): ordered sequence of interpoint distances.

Guess: number of distances smaller than r approximately Poisson.

Let I_{ij} (number of points randomly!) indicate whether distance between i^{th} and j^{th} points is smaller than r . To couple, fix pair ij (Take care about conditioning here!). Construct U_{ij} directly by summing short distances from the process. Construct V_{ij} by conditioning on $I_{ij} = 1$ and omitting distance ij : use Palm representation of conditioning on location of points i, j to obtain $V_{ij} \geq U_{ij}$. Setting $\lambda = \mathbb{E}[W]$, deduce

$$\text{dist}_{\text{TV}}(\mathcal{L}(W), \text{Poisson}(\lambda)) \leq (1 - \exp(-\lambda)) (\lambda^{-1} \mathbf{Var}[W] - 1). \quad (9)$$

Remaining integrations are exercise for reader :-).

5 Mixing

How fast does a Markov chain converge towards equilibrium? Non-coupling methods are very important, but coupling can also be useful.

5.1 Coupling inequality

Suppose X is a Markov chain, with equilibrium distribution π , for which we can produce a coupling between any two points x, y , which succeeds at time $T_{x,y} < \infty$. Then

$$\text{dist}_{tv}(\mathcal{L}(X_n), \pi) \leq \max_y \{\mathbb{P}[T_{x,y} > t]\}. \quad (10)$$

Useful bounds depend on finding and analyzing the right coupling!

This is also the basis for an empirical approach (Johnson 1996).

5.2 Strong stationary time

Definition 2 (*Aldous and Diaconis 1987; Diaconis and Fill 1990*) The random stopping time T is a strong stationary time for the Markov chain X (whose equilibrium distribution is π) if

$$\mathbb{P}[T = k, X_k = s] = \pi_s \times \mathbb{P}[T = k].$$

Application to card-shuffling (*Broder*): notion of *checked* cards. Transpose by choosing 2 cards at random; LH, RH. Check card pointed to by LH if

either LH, RH are the same unchecked card;
or LH is unchecked, RH is checked.

Inductive claim: given number of checked cards, positions in pack of checked cards, *list* of values of cards;

the map of checked card to value is uniformly random.

Let T_m be time when m cards checked. Then (induction) $T = \sum_{m=0}^{n-1} T_m$ is a strong stationary time.

How big is T ?

We know $T_{m+1} - T_m$ is Geometrically distributed with success probability $(n - m)(m + 1)/n^2$. So mean of T is

$$\mathbb{E} \left[\sum_{m=0}^{n-1} T_m \right] = \sum_{m=0}^{n-1} \frac{n^2}{n+1} \left(\frac{1}{n-m} + \frac{1}{m+1} \right) \approx 2n \log n.$$

MOREOVER: $T = \sum_{m=0}^{n-1} T_m$ is a discrete version of time to complete infection for simple epidemic (without recovery) in n individuals, starting with 1 infective, individuals contacting at rate n^2 . A classic calculation tells us $T/(2n \log n)$ has a limiting distribution.

NOTE: by group representation theory the correct asymptotic is $\frac{1}{2}n \log n$.

6 The Coupling zoo

independent (Doebelin)
reflection
synchronized
Ornstein (random walks)

Vasershtein

maximal (non-adaptive)
(Griffeath 1975)
(Goldstein 1979)
time-changed

shift
(Aldous and Thorisson 1993)

Before coupling ...

Be **independent**.

Do the **opposite**.

Do the **same**.

Be a **liberal democrat**:

agree on big steps,
otherwise be independent.

Take each step so as to

maximize immediate success.

Be an **inside-trader**:

maximize success over each
initial segment of steps!

Relax success criterion:

slow one chain, **speed up** other.

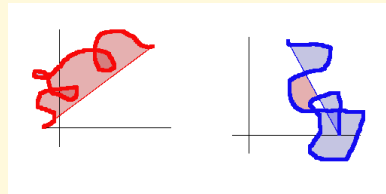
Relax success criterion:

aim to do the same thing
but at **different times!**

7 Exotics

7.1 Planar Brownian motion and Lévy area

- Hypoelliptic diffusion $(A, B, \int A dB - B dA)$;
- $\int A dB - B dA$ is “signed area” swept out by segment from origin to Brownian point (A, B) ;
- Couple *via* controls using “Liapunov function” ideas: need reflection *and* synchronized coupling – and “twisted” coupling helps too! (Ben Arous et al. 1995)



7.2 Real Brownian motion and iterated time integrals

How far can we go?

B alone

Use reflection (Lindvall 1983)

$(B, \int B dt)$

Also need synchronized (Ben Arous et al. 1995)

$(B, \int B dt, \int \int B ds dt)$

(Kendall and Price 2003)

$(B, \int B dt, \int \int B ds dt, \int \int \int B du ds dt, \dots)$


can we couple finitely many?

We certainly can't couple all (use *eg* Laplace transforms).




We *can* couple finitely many!

References

Aldous, D. J. and P. Diaconis [1987]. Strong uniform times and finite random walks. *Advances in Applied Mathematics* 8(1), 69–97.

Aldous, D. J. and J. A. Fill [200x]. Reversible Markov chains and random walks on graphs. Book in preparation .

Aldous, D. J. and H. Thorisson [1993]. Shift-coupling. *Stochastic Processes and Their Applications* 44(1), 1–14. [doi>](#)

This is a rich hypertext bibliography. Journals are linked to their homepages, and icons  which link to preprints are inserted where available. Stable URL links (as provided for example by JSTOR  or Project Euclid ) have been added where known. Access to such URLs is not universal: in case of difficulty you should check whether you are registered (directly or indirectly) with the relevant provider.

Athreya, K. B. and P. Ney [1978]. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society* 245, 493–501.

Ball, F. and P. Donnelly [1995]. Strong approximations for epidemic models. *Stochastic Processes and Their Applications* 55(1), 1–21. [doi>](#)

Barbour, A. D., L. Holst, and S. Janson [1992]. *Poisson approximation*, Volume 2 of *Oxford Studies in Probability*. New York: The Clarendon Press Oxford University Press. Oxford Science Publications.

Bass, R. F. and K. Burdzy [1993]. On domain monotonicity of the Neumann heat kernel. *Journal of Functional Analysis* 116(1), 215–224. [doi>](#)

Ben Arous, G., M. Cranston, and W. S. Kendall [1995]. Coupling constructions for hypoelliptic diffusions: Two examples. In M. Cranston and M. Pinsky (Eds.), *Stochastic Analysis: Summer Research*

[Home Page](#)[Title Page](#)

Page 25 of 27


[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Institute July 11-30, 1993., Volume 57, Providence, RI Providence, pp. 193–212. [American Mathematical Society](#).

Breiman, L. [1992]. *Probability*, Volume 7 of *Classics in Applied Mathematics*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). Corrected reprint of the 1968 original.

Carmona, R. A. and W. A. Zheng [1994]. Reflecting Brownian motions and comparison theorems for Neumann heat kernels. *Journal of Functional Analysis* 123(1), 109–128. [doi>](#)

Chavel, I. [1986]. Heat diffusion in insulated convex domains. *The Journal of the London Mathematical Society (Second Series)* 34(3), 473–478.

Diaconis, P. and J. Fill [1990]. Strong stationary times via a new form of duality. *The Annals of Probability* 18, 1483–1522. 

Doebelin, W. [1938]. Exposé de la théorie des chaînes simples constants de

Markoff á un nombre fini d'états. *Revue Math. de l'Union Interbalkanique* 2, 77–105.

Goldstein, S. [1978 / 1979]. Maximal coupling. *Zeitschrift für Wahrscheinlichkeitstheorie und Verve Gebiete* 46(2), 193–204.

Griffeath, D. [1974 / 1975]. A maximal coupling for Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verve Gebiete* 31, 95–106.

Hägström, O. and R. Pemantle [1998]. First passage percolation and a model for competing spatial growth. *Journal of Applied Probability* 35(3), 683–692. [doi>](#)

Hägström, O. and R. Pemantle [2000]. Absence of mutual unbounded growth for almost all parameter values in the two-type Richardson model. *Stochastic Processes and Their Applications* 90(2), 207–222. [doi>](#)

[Home Page](#)[Title Page](#)

Page 26 of 27

[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Hsu, E. P. [1994]. A domain monotonicity property of the Neumann heat kernel. *Osaka J. Math.* 31(1), 215–223.

Johnson, V. E. [1996]. Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths. *Journal of the American Statistical Association* 91(433), 154–166.

Kendall, W. S. [1989]. Coupled Brownian motions and partial domain monotonicity for the Neumann heat kernel. *Journal of Functional Analysis* 86, 226–236.

Kendall, W. S. and G. Montana [2002, May]. Small sets and Markov transition densities. *Stochastic Processes and Their Applications* 99(2), 177–194. also University of Warwick Department of Statistics Research Report 371

Kendall, W. S. and C. J. Price [2003]. Coupling the twice-iterated Kolmogorov diffusion. In preparation.

Kendall, W. S. and I. W. Saunders [1983]. Epidemics in competition II: The general epidemic. *Journal of the Royal Statistical Society (Series B: Methodological)* 45, 238–244.

Lindvall, T. [1983]. On coupling of diffusion processes. *Journal of Applied Probability* 20(1), 82–93.

Lindvall, T. [1991]. W. Doeblin 1915–1940. *The Annals of Probability* 19(3), 929–934.

Lindvall, T. [2002]. *Lectures on the coupling method*. Mineola, NY: Dover Publications Inc. Corrected reprint of the 1992 original.

Meyn, S. P. and R. L. Tweedie [1993]. *Markov Chains and Stochastic Stability*. New York: Springer-Verlag.

Mollison, D. [1977]. Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society (Series B: Methodological)* 39(3), 283–326.

[Introduction](#)[Monotonicity](#)[Representation](#)[Approximation](#)[Mixing](#)[The Coupling Zoo](#)[Exotics](#)[Home Page](#)[Title Page](#)[Page 27 of 27](#)[Go Back](#)[Full Screen](#)[Close](#)[Quit](#)

Nummelin, E. [1978]. A splitting technique for Harris-recurrent chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwe Gebiete* 43, 309–318.

Roberts, G. O. and J. S. Rosenthal [2001]. Small and pseudo-small sets for Markov chains. *Stochastic Models* 17(2), 121–145.

Silverman, B. W. and T. C. Brown [1978]. Short distances, flat triangles and Poisson limits. *Journal of Applied Probability* 15(4), 815–825.

Thorisson, H. [2000]. *Coupling, stationarity, and regeneration*. New York: Springer-Verlag.