Lectures on Coupling
EPSRC/RSS GTP course, September 2005

Wilfrid Kendall
w.s.kendall@warwick.ac.uk

Department of Statistics, University of Warwick

12th–16th September 2005
<table>
<thead>
<tr>
<th>Lectures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>Coupling and Monotonicity</td>
</tr>
<tr>
<td>Representation</td>
</tr>
<tr>
<td>Approximation using coupling</td>
</tr>
<tr>
<td>Mixing of Markov chains</td>
</tr>
<tr>
<td>Conclusion</td>
</tr>
</tbody>
</table>
Monday 12 September
09.00–09.05 Welcome
09.05–10.00 Coupling 1 Introduction
10.00–11.00 Bayes 1
11.00–11.30 Coffee/Tea
11.30–12.30 Bayes 2
12.30–13.45 Lunch
13.45–14.45 Coupling 2 Monotonicity
14.45–15.45 Coupling 3 Representation
15.45–16.15 Coffee/Tea
16.15–17.15 Bayes 3
19.00–21.00 Dinner

Tuesday 13 September
09.00–10.00 Bayes 4
10.00–11.00 Bayes Practical 1
11.00–11.30 Coffee/Tea
11.30–12.30 Coupling 4 Approximation
12.30–13.45 Lunch
13.45–14.45 Coupling 5 Mixing
14.45–15.45 Coupling practical A0.02
15.45–16.15 Coffee/Tea
16.15–17.15 Bayes 5
19.00–21.00 Dinner

Wednesday 14 September
09.00–10.00 Coupling 6 The Zoo
10.00–11.00 Coupling 7 Perfection: CFTP (I)
11.00–11.30 Coffee/Tea
11.30–12.30 Bayes Exercises
12.30–13.45 Lunch
13.45–18.00 Excursion
19.00–21.00 Dinner

Thursday 15 September
09.00–10.00 Bayes 6
10.00–11.00 Bayes 7
11.00–11.30 Coffee/Tea
11.30–12.30 Coupling practical A0.02
12.30–13.45 Lunch
13.45–14.45 Coupling 8 Perfection: CFTP (II)
14.45–15.45 Coupling 9 Perfection: FMMR
15.45–16.15 Coffee/Tea
16.15–17.15 Bayes 8
19.00–21.00 Dinner (Sutherland Suite)

Friday 16 September
09.00–10.00 Coupling 10 Sundry topics
10.00–11.00 Bayes Practical 2 (MS.04/A1.01)
11.00–11.30 Coffee/Tea
11.30–13.00 Bayes Practical 2 (MS.04/A1.01)
13.00 Lunch/departure

(all lectures in MS.04)
“The best thing for being sad,” replied Merlin, beginning to puff and blow, “is to learn something. That’s the only thing that never fails. You may grow old and trembling in your anatomies, you may lie awake at night listening to the disorder of your veins, you may miss your only love, you may see the world about you devastated by evil lunatics, or know your honour trampled in the sewers of baser minds. There is only one thing for it then – to learn. Learn why the world wags and what wags it. That is the only thing which the mind can never exhaust, never alienate, never be tortured by, never fear or distrust, and never dream of regretting. Learning is the only thing for you. Look what a lot of things there are to learn.”

— T. H. White, “The Once and Future King”
A brief description of coupling

“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 $Y$, leads to clear intuitive explanations of important facts about $X$. 
Aim and style of lectures

These lectures aim to survey ideas from coupling theory, using a pattern of beginning with an intuitive example, developing the idea of the example, and then remarking on further ramifications of the theory.

Carelessly planned projects take three times longer to complete than expected. Carefully planned projects take four times longer to complete than expected, mostly because the planners expect their planning to reduce the time it takes.
Aim and style of lectures

These lectures aim to survey ideas from coupling theory, using a pattern of beginning with an intuitive example, developing the idea of the example, and then remarking on further ramifications of the theory.

Aiming for style of Rubeus Hagrid’s “Care of Magical Creatures” rather than Dolores Umbridge’s “Defence against the Dark Arts”.

*Carelessly planned projects take three times longer to complete than expected. Carefully planned projects take four times longer to complete than expected, mostly because the planners expect their planning to reduce the time it takes.*
Reading and browsing
Books and URIs

- History: Doeblin (1938), see also Lindvall (1991).

Of making many books there is no end, and much study wearies the body.
— Ecclesiastics 12:12b
Reading and browsing

Books and URIs

- History: Doeblin (1938), see also Lindvall (1991).

Of making many books there is no end, and much study wearies the body.
— Ecclesiastics 12:12b
Reading and browsing
Books and URIs

- History: Doeblin (1938), see also Lindvall (1991).

http://www.warwick.ac.uk/go/wsk/talks/gtp.pdf
http://research.microsoft.com/~dbwilson/exact/

Of making many books there is no end,
and much study wearies the body.
— Ecclesiastics 12:12b
Soundbites

*Probability theory has a right and a left hand*  
— *Breiman (1992, Preface).*
Soundbites

Probability theory has a right and a left hand

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

*Probability theory has a right and a left hand*

*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.*
Soundbites

*Probability theory has a right and a left hand*  

*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.*
Probability theory has a right and a left hand  

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.

Coupling is the soul of probability.
Card shuffling
Top card shuffle

**Draw:** Top card moved to random location in pack.

**Q:** How long to equilibrium?

▶ **Hint:** consider the various possible orders of the cards lying below the specified card at each stage.

▶ Note existence of special states, such that you should stop after moving from such a state (unusual in general!).

▶ Mean time till equilibrium is order $n \log(n)$. 
**Card shuffling**

**Top card shuffle**

**Draw:** Top card moved to random location in pack.

**Q:** How long to equilibrium?

**A:** Two alternative strategies:

(a) wait till **bottom card** gets to top, then draw one more;
**Draw:** Top card moved to random location in pack.

**Q:** How long to equilibrium?

**A:** Two alternative strategies:

(a) wait till **bottom card** gets to top, then draw one more;

(b) or wait till **next-to-bottom card** gets to top, then draw one more.

**HINT:** Consider the various possible orders of the cards lying below the specified card at each stage.

**Note** existence of special states, such that you should stop after moving from such a state (unusual in general!).

**Mean time** till equilibrium is order $n \log(n)$.
Card shuffling

Top card shuffle

**Draw:** Top card moved to random location in pack.

**Q:** How long to equilibrium?

**A:** Two alternative strategies:

(a) wait till **bottom card** gets to top, then draw one more;

(b) or wait till **next-to-bottom card** gets to top, then draw one more.

▶ **HINT:** consider the various possible orders of the cards lying below the **specified card** at each stage.

▶ Note existence of special states, such that you should stop after moving from such a state (unusual in general!).

▶ Mean time till equilibrium is order \( n \log(n) \).
**Card shuffling**

**Top card shuffle**

**Draw:** Top card moved to random location in pack.

**Q:** How long to equilibrium?

**A:** Two alternative strategies:

(a) wait till **bottom card** gets to top, then draw one more;

(b) or wait till **next-to-bottom card** gets to top, then draw one more.

▶ **HINT:** consider the various possible orders of the cards lying below the **specified card** at each stage.

▶ Note existence of special states, such that you should stop after moving from such a state (unusual in general!).

**Mean time till equilibrium is order** $n \log n$. 
Card shuffling
Top card shuffle

**Draw:** Top card moved to random location in pack.

**Q:** How long to equilibrium?

**A:** Two alternative strategies:

(a) wait till **bottom card** gets to top, then draw one more;

(b) or wait till **next-to-bottom card** gets to top, then draw one more.

▶ **HINT:** consider the various possible orders of the cards lying below the **specified card** at each stage.

▶ Note existence of special states, such that you should stop after moving from such a state (unusual in general!).

▶ Mean time till equilibrium is order $n \log(n)$. 

▶ **EXERCISE 1.1**
Card shuffling
Riffle shuffle

**Draw:** split card pack into two parts using Binomial distribution. Recombine uniformly at random.

**Q:** How long to equilibrium (uniform randomness)?
Card shuffling
Riffle shuffle

**Draw:** split card pack into two parts using Binomial distribution. Recombine uniformly at random.

**Q:** How long to equilibrium (uniform randomness)?

▶ **DEVIous HINT:** Apply a time reversal and re-labelling, so reversed process looks like this: to each card assign a sequence of random bits (0, 1). Remove cards with top bit set, move to top of pack, remove top bits, repeat . . . .
Card shuffling
Riffle shuffle

**Draw:** split card pack into two parts using Binomial distribution. Recombine uniformly at random.

**Q:** How long to equilibrium (uniform randomness)?

- **DEVIOUS HINT:** Apply a time reversal and re-labelling, so reversed process looks like this: to each card assign a sequence of random bits (0, 1). Remove cards with top bit set, move to top of pack, remove top bits, repeat . . . .

- How long till resulting random permutation is uniform? Depends on how long sequence must be in order to label each card uniquely.
Card shuffling
Riffle shuffle

**Draw:** split card pack into two parts using Binomial distribution. Recombine uniformly at random.

**Q:** How long to equilibrium (uniform randomness)?

▶ **DEVIOUS HINT:** Apply a time reversal and re-labelling, so reversed process looks like this: to each card assign a sequence of random bits (0, 1). Remove cards with top bit set, move to top of pack, remove top bits, repeat . . . .

▶ How long till resulting random permutation is uniform? Depends on how long sequence must be in order to label each card uniquely.

▶ Time-reversals will be important later when discussing queues, dominated *CFTP*, and Siegmund duality.
Historical example
First appearance of coupling

Treatment of convergence of Markov chains to statistical equilibrium by Doeblin (1938).

Theorem 1.1 (Doeblin coupling)

For a finite state space Markov chain, consider two copies; one started in equilibrium ($\mathbb{P} [X = j] = \pi_j$), one at some specified starting state. Run the chains independently till they meet, or couple. Then:

$$\frac{1}{2} \sum_j |\pi_j - p_{ij}^{(n)}| \leq \mathbb{P} [\text{no coupling by time } n] \quad (1)$$

We develop the theory for this later, when we discuss the coupling inequality.
Historical example
Successful coupling

Definition 1.2
We say a coupling between two random processes $X$ and $Y$ succeeds if almost surely $X$ and $Y$ eventually meet.
Definition 1.2
We say a coupling between two random processes $X$ and $Y$ succeeds if almost surely $X$ and $Y$ eventually meet.

The Doeblin coupling succeeds for all finite aperiodic irreducible Markov chains (via the lemma which says there is $T \geq 0$ such that $p_{ij}^{(n)} > 0$ once $n \geq N$), and indeed for all countable positive-recurrent aperiodic Markov chains.
Definition 1.2
We say a coupling between two random processes $X$ and $Y$ \textit{succeeds} if almost surely $X$ and $Y$ eventually meet.

The Doeblin coupling \textit{succeeds} for all finite aperiodic irreducible Markov chains \textit{(via} the lemma which says there is $T \geq 0$ such that $p_{ij}^{(n)} > 0$ once $n \geq N$), and indeed for all countable \textit{positive-recurrent} aperiodic Markov chains. Successful couplings control rate of convergence to equilibrium.
Definition 1.2
We say a coupling between two random processes $X$ and $Y$ succeeds if almost surely $X$ and $Y$ eventually meet.

The Doeblin coupling succeeds for all finite aperiodic irreducible Markov chains (via the lemma which says there is $T \geq 0$ such that $p_{ij}^{(n)} > 0$ once $n \geq N$), and indeed for all countable positive-recurrent aperiodic Markov chains. Successful couplings control rate of convergence to equilibrium.

Thorisson has constructed a coupling proof that $p_{ij}^{(n)} \to 0$ for countable null-recurrent Markov chains.
Coupling and Monotonicity

MONOTONOUS

ADJECTIVE: Arousing no interest or curiosity: boring, drear, dreary, dry, dull, humdrum, irksome, stuffy, tedious, tiresome, uninteresting, weariful, wearisome, weary. See EXCITE.


Coupling and Monotonicity
Binomial monotonicity
Rabbits
FKG inequality
Fortuin-Kasteleyn representation
Binomial monotonicity

The simplest examples of monotonicity arise for Binomial random variables, thought of as sums of Bernoulli random variables.

- Suppose $X$ is distributed as Binomial$(n, p)$. Show that $\mathbb{P}[X \geq k]$ is an increasing function of the success probability $p$. 

**EXERCISE 2.1**

Generalize to show uniqueness of critical probability for percolation on a Euclidean lattice.
Binomial monotonicity

The simplest examples of monotonicity arise for Binomial random variables, thought of as sums of Bernoulli random variables.

- Suppose $X$ is distributed as $\text{Binomial}(n, p)$. Show that $\mathbb{P}[X \geq k]$ is an increasing function of the success probability $p$.

- Generalize to show uniqueness of critical probability for percolation on a Euclidean lattice.
Increasing events

Here is the notion which makes these examples work.

Definition 2.1
Consider a sequence of (binary) random variables $Y_1, Y_2, \ldots, Y_n$. An *increasing event* $A$ for this sequence is determined by the values $Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n$, and the indicator $\mathbb{I}[A]$ is an increasing function of the $y_1, y_2, \ldots, y_n$. 
Increasing events

Here is the notion which makes these examples work.

**Definition 2.1**
Consider a sequence of (binary) random variables $Y_1, Y_2, \ldots, Y_n$. An *increasing event* $A$ for this sequence is determined by the values $Y_1 = y_1$, $Y_2 = y_2$, $\ldots$, $Y_n = y_n$, and the indicator $I[A]$ is an increasing function of the $y_1, y_2, \ldots, y_n$.

Both $[X \geq k]$ and “there is an infinite connected component” are increasing events. The idea is developed further in the discussion of the FKG inequality below.
Continuous Rabbits

Kendall and Saunders (1983) used coupling to analyze competing myxomatosis epidemics in Australian rabbits.

\[ i_1' = \alpha_1 \beta_1 s_i - \beta_1 i_1, \]

(2)

Infectives \( i_1 \),
Continuous Rabbits

Kendall and Saunders (1983) used coupling to analyze competing myxomatosis epidemics in Australian rabbits.

\[
\begin{align*}
i_1' &= \alpha_1 \beta_1 s i_1 - \beta_1 i_1, \\
i_2' &= \alpha_2 \beta_2 s i_2 - \beta_2 i_2, \\
\end{align*}
\]

(2)

Infectives \(i_1, i_2\);
Continuous Rabbits

Kendall and Saunders (1983) used coupling to analyze competing myxomatosis epidemics in Australian rabbits.

\begin{align*}
  i'_1 &= \alpha_1 \beta_1 s_i - \beta_1 i_1, \\
  i'_2 &= \alpha_2 \beta_2 s_i - \beta_2 i_2,
\end{align*}

(2)

Suppose \( \alpha_1 > \alpha_2 \).

Infectives \( i_1, i_2 \);
Continuous Rabbits

Kendall and Saunders (1983) used coupling to analyze competing myxomatosis epidemics in Australian rabbits.

\[
\begin{align*}
    s' &= -\alpha_1 \beta_1 s_i - \alpha_2 \beta_2 s_i \\
    i'_1 &= \alpha_1 \beta_1 s_i - \beta_1 i_1 \\
    i'_2 &= \alpha_2 \beta_2 s_i - \beta_2 i_2,
\end{align*}
\]

(2)

Suppose \( \alpha_1 > \alpha_2 \).

Susceptibles \( s \); Infectives \( i_1, i_2 \);
Continuous Rabbits

Kendall and Saunders (1983) used coupling to analyze competing myxomatosis epidemics in Australian rabbits.

\[ s' = -\alpha_1 \beta_1 s_i - \alpha_2 \beta_2 s_i \]
\[ i'_1 = \alpha_1 \beta_1 s_i - \beta_1 i_1, \quad r'_1 = \beta_1 i_1 \]
\[ i'_2 = \alpha_2 \beta_2 s_i - \beta_2 i_2, \quad r'_2 = \beta_2 i_2 \] (2)

Suppose \( \alpha_1 > \alpha_2 \).

Susceptibles \( s \); Infectives \( i_1, i_2 \); Removals \( r_1, r_2 \).
Continuous Rabbits

Kendall and Saunders (1983) used coupling to analyze competing myxomatosis epidemics in Australian rabbits.

\[ s' = -\alpha_1 \beta_1 s_i_1 - \alpha_2 \beta_2 s_i_2 \]
\[ i'_1 = \alpha_1 \beta_1 s_i_1 - \beta_1 i_1 , \quad r'_1 = \beta_1 i_1 \]
\[ i'_2 = \alpha_2 \beta_2 s_i_2 - \beta_2 i_2 , \quad r'_2 = \beta_2 i_2 \quad (2) \]

Suppose \( \alpha_1 > \alpha_2 \).

Susceptibles \( s \); Infectives \( i_1, i_2 \); Removals \( r_1, r_2 \).

Are \( r_1(\infty), r_2(\infty) \) appropriately monotonic in \( i_1, i_2 \)?
Continuous Rabbits

Kendall and Saunders (1983) used coupling to analyze competing myxomatosis epidemics in Australian rabbits.

\[ s' = -\alpha_1 \beta_1 s_i - \alpha_2 \beta_2 s_i \]
\[ i'_1 = \alpha_1 \beta_1 s_i - \beta_1 i_1 , \quad r'_1 = \beta_1 i_1 \]
\[ i'_2 = \alpha_2 \beta_2 s_i - \beta_2 i_2 , \quad r'_2 = \beta_2 i_2 \]  

(2)

Suppose \( \alpha_1 > \alpha_2 \).

Susceptibles \( s \); Infectives \( i_1, i_2 \); Removals \( r_1, r_2 \).

Are \( r_1(\infty), r_2(\infty) \) appropriately monotonic in \( i_1, i_2 \)?
Stochastic Rabbits

Trick: the right stochastic model has required monotonicity!
Stochastic Rabbits

Trick: the right stochastic model has required monotonicity!

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

Trick: the right stochastic model has required monotonicity!

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

Trick: the right stochastic model has required monotonicity!

- **Stochastic model.** List potential infections from each individual as times from infection (*nb:* different rates for **type-1** and **type-2**);
Stochastic Rabbit Trick: the right stochastic model has required monotonicity!

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

Converting type-1 initial infective to susceptible or to type-2 infective "clearly" delays type-1 epidemic: hence desired monotonicity for stochastic model.

Stochastic Rabbits

Trick: the right stochastic model has required monotonicity!

▶ Stochastic model. List potential infections from each individual as times from infection (\textit{nb:} different rates for \textbf{type-1} and \textbf{type-2});
Stochastic Rabbits

Trick: the right stochastic model has required monotonicity!

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

Trick: the right stochastic model has required monotonicity!

- **Stochastic model.** List potential infections from each individual as times from infection (*nb:* different rates for **type-1** and **type-2**);
- Converting type-1 initial infective to susceptible or to type-2 infective “clearly” delays type-1 epidemic: hence desired monotonicity for stochastic model.

Deterministic model inherits monotonicity.

Just one out of many applications to epidemic theory: see also Ball and Donnelly (1995).

For coupling in spatial epidemics, see Mollison (1977), Häggström and Pemantle (1998, 2000).
Stochastic Rabbits

Trick: the right stochastic model has required monotonicity!

- **Stochastic model.** List potential infections from each individual as times from infection (*nb:* different rates for **type-1** and **type-2**);
- Converting type-1 initial infective to susceptible or to type-2 infective “clearly” delays type-1 epidemic: hence desired monotonicity for stochastic model.
- Deterministic model inherits monotonicity.
Stochastic Rabbits

Trick: the right stochastic model has required monotonicity!

- **Stochastic model.** List potential infections from each individual as times from infection (*nb:* different rates for type-1 and type-2);
- Converting type-1 initial infective to susceptible or to type-2 infective “clearly” delays type-1 epidemic: hence desired monotonicity for stochastic model.
- Deterministic model inherits monotonicity.

FKG inequality

Theorem 2.2 (FKG for independent Bernoulli set-up)
Suppose $A, B$ are two increasing events for binary $Y_1, Y_2, \ldots, Y_n$: then they are positively correlated;

\[ P[A \cap B] \geq P[A]P[B]. \quad (3) \]
Theorem 2.2 (FKG for independent Bernoulli set-up)

Suppose $A$, $B$ are two increasing events for binary $Y_1$, $Y_2$, ..., $Y_n$: then they are positively correlated;

$$P[A \cap B] \geq P[A]P[B].$$

(3)

The FKG inequality generalizes:

- replace sets $A$, $B$ by “increasing random variables” (Grimmett 1999, §2.2);
- allow $Y_1$, $Y_2$, ..., $Y_n$ to “interact attractively” (Preston 1977, or vary as described in the Exercise!).

EXERCISE 2.4
Fortuin-Kasteleyn representation

*Site-percolation* can be varied to produce *bond-percolation*: for a given graph $G$ suppose the edges or *bonds* are independently open or not to the flow of a fluid with probability $p$. The probability of any given configuration is

$$p^{\#(\text{open bonds})} \times (1 - p)^{\#(\text{closed bonds})}. \quad (4)$$

Suppose we want a new *dependent* bond-percolation model, biased towards the formation of many different connected components or clusters in the resulting random graph. **Definition 2.3 (Random Cluster model)**

The probability of any given configuration is proportional to

$$q^{\#(\text{components})} \times p^{\#(\text{open bonds})} \times (1 - p)^{\#(\text{closed bonds})}. \quad (5)$$
Fortuin-Kasteleyn representation

*Site-percolation* can be varied to produce *bond-percolation*: for a given graph $G$ suppose the edges or *bonds* are independently open or not to the flow of a fluid with probability $p$. The probability of any given configuration is

$$p^{\#(\text{open bonds})} \times (1 - p)^{\#(\text{closed bonds})}.$$  \hspace{1cm} (4)

Suppose we want a new *dependent* bond-percolation model, biased towards the formation of many different connected components or clusters in the resulting random graph.
**Fortuin-Kasteleyn representation**

*Site-percolation* can be varied to produce *bond-percolation*: for a given graph $G$ suppose the edges or *bonds* are independently open or not to the flow of a fluid with probability $p$. The probability of any given configuration is

$$p^\#(\text{open bonds}) \times (1 - p)^\#(\text{closed bonds}).$$

(4)

Suppose we want a new *dependent* bond-percolation model, biased towards the formation of many different connected components or clusters in the resulting random graph.

**Definition 2.3 (Random Cluster model)**

The probability of any given configuration is *proportional to*

$$q^\#(\text{components}) \times p^\#(\text{open bonds}) \times (1 - p)^\#(\text{closed bonds}).$$

(5)
Suppose $q = 2$ and we assign signs $\pm 1$ independently to each of the resulting clusters. This produces a random configuration assigning $S_i = \pm 1$ to each site $i$ (according to which cluster contains it): a spin model. Remarkably:

**Theorem 2.4 (Fortuin-Kasteleyn representation)**

The spin model described above is the Ising model on $G$: the probability of any given configuration is proportional to

$$
\exp \left( \frac{1}{2} \beta \sum_{i \sim j} S_i S_j \right) \propto \exp (\beta \times \#(\text{similar pairs})) \quad (6)
$$

where $p = 1 - e^{-\beta}$.
Fortuin-Kastelyn representation (ctd.)

Suppose $q = 2$ and we assign signs $\pm 1$ independently to each of the resulting clusters. This produces a random configuration assigning $S_i = \pm 1$ to each site $i$ (according to which cluster contains it): a *spin model*. Remarkably:

**Theorem 2.4 (Fortuin-Kasteleyn representation)**

*The spin model described above is the Ising model on $G$: the probability of any given configuration is proportional to*

$$\exp \left( \frac{1}{2} \beta \sum_{i \sim j} S_i S_j \right) \propto \exp (\beta \times \#(\text{similar pairs})) \quad (6)$$

*where $p = 1 - e^{-\beta}$.*

We have **coupled** the Ising model (of statistical mechanics and image analysis fame) to a dependent bond-percolation model!
Moreover, we can compare probabilities of increasing events for bond-percolation and random cluster models:

**Theorem 2.5 (Fortuin-Kasteleyn comparison)**

Suppose $A$ is an increasing event for bond-percolation. Then

$$\mathbb{P} [A \text{ under bond percolation}(p)] \leq \mathbb{P} [A \text{ under random cluster}(p', q)] \leq \mathbb{P} [A \text{ under bond percolation}(p')]$$

when $q \geq 1$ and $p/(1 - p) = p'/(q(1 - p'))$.

The FKG inequality holds for the general random cluster model when $q \geq 1$. 

“The good Christian should beware of mathematicians and all those who make empty prophecies. The danger already exists that mathematicians have made a covenant with the devil to darken the spirit and confine man in the bonds of Hell.”

— St. Augustine
Wrights’ Axioms of Queuing Theory.

1. *If you have a choice between waiting here and waiting there, wait there.*
2. *All things being equal, it is better to wait at the front of the line.*
3. *There is no point in waiting at the end of the line.*
   
   **But note Wrights’ Paradox:**
4. *If you don’t wait at the end of the line, you’ll never get to the front.*
5. *Whichever line you are in, the others always move faster.*

— Charles R. B. Wright

Representation

Queues
Strassen’s result
Princesses and Frogs
Age brings avarice
Capitalist geography
Back to probability
Split chains and small sets
Queues

We have already seen how coupling can provide important representations: the Fortuin-Kastelyn representation for the Ising model. Here is another representation, from the theory of queues, which also introduces ideas of time reversal which will be important later.
Queues

We have already seen how coupling can provide important representations: the Fortuin-Kastelyn representation for the Ising model. Here is another representation, from the theory of queues, which also introduces ideas of time reversal which will be important later.

Recall the standard notation for queues: an $M/M/1$ queue has Markov inputs (arrivals according to Poisson process), Markov service times (Exponential, which counts as Markov by memoryless property), and $1$ server.
Queues

We have already seen how coupling can provide important representations: the Fortuin-Kastelyn representation for the Ising model. Here is another representation, from the theory of queues, which also introduces ideas of time reversal which will be important later.

Recall the standard notation for queues: an $M/M/1$ queue has Markov inputs (arrivals according to Poisson process), Markov service times (Exponential, which counts as Markov by memoryless property), and 1 server.

When we move away from Markov then analysis gets harder: if inputs are GI (General Input, but independent) or if service times are General (but still independent) then we can use embedding methods to reduce to discrete time Markov chain theory. If both ($GI/G/1$, independent service and inter-arrival times) then even this is not available!
Lindley’s representation I

However Lindley noticed a beautiful representation for waiting time $W_n$ of customer $n$ in terms of services $S_n$ and interarrivals $X_n$...

**Theorem 3.1 (Lindley’s equation)**

Consider the GI/G/1 queue waiting time identity.

$$W_{n+1} = \max \{0, W_n + S_n - X_{n+1}\} = \max \{0, W_n + \eta_n\}$$

$$= \max \{0, \eta_n, \eta_n + \eta_{n-1}, \ldots, \eta_n + \eta_{n-1} + \ldots + \eta_1\}$$

$$= \mathcal{D} \max \{0, \eta_1, \eta_1 + \eta_2, \ldots, \eta_1 + \eta_2 + \ldots + \eta_n\}$$

and thus we obtain the steady-state expression

$$W_\infty = \mathcal{D} \max \{0, \eta_1, \eta_1 + \eta_2, \ldots\} . \quad (7)$$

If $\text{Var} [\eta_i] < \infty$ then SLLN/CLT/random walk theory shows $W_\infty$ will be finite if and only if $\mathbb{E} [\eta_i] < 0$ or $\eta_i \equiv 0$. 
Coupling enters in at the crucial time-reversal step! This idea will reappear later in our discussion of the falling-leaves model . . . .
Supposing we lose independence? Loynes (1962) discovered a coupling application to queues with (for example) general dependent stationary inputs and associated service times, pre-figuring CFTP.
Theorem 3.2
Suppose queue arrivals follow a stationary point process stretching back to time $-\infty$. Denote arrivals/associated service times in $(s, t]$ by $N_{s,t}$ (stationarity: statistics of process $\{N_{s,s+u}: u \geq 0\}$ do not depend on $s$). Let $Q^T$ denote the behaviour of the queue observed from time 0 onwards if begun with 0 customers at time $-T$. The queue converges to statistical equilibrium if and only if

$$\lim_{T \to \infty} Q^T \text{ exists almost surely.}$$

Strassen’s result

When can we closely couple\(^1\) two random variables? This question is easy to deal with in one dimension, using the inverse probability transform, taken up below. However there is also a beautiful treatment of Strassen’s treatment of the multivariate case, based on the Marriage Lemma and due to Dudley (1976). Here is a fairy-tale explanation, building on Pollard (2001)’s charming exposition.

\(^1\)Relates to the \textit{Prohorov metric}. 
Princesses and Frogs

“You gotta kiss a thousand frogs before you find a prince!”
Princesses and Frogs

“You gotta kiss a thousand frogs before you find a prince!”

- World $\mathcal{X}$ contains finite supply $S$ of princesses and $F$ of frogs.
Princesses and Frogs

“You gotta kiss a thousand frogs before you find a prince!”

- World $\mathcal{X}$ contains finite supply $S$ of princesses and $F$ of frogs.
- Each princess $\sigma$ has list $L(\sigma) \subseteq F$ of eligible frogs.

Lemma 3.3 (Marriage Lemma)

A suitable injection $f: S \rightarrow F$ exists exactly when for all sets of princesses $A \subseteq S$, $\#A \leq \#L(A)$ for $L(A) = \bigcup \{L(\sigma) : \sigma \in A\}$. (8)
Princesses and Frogs

“You gotta kiss a thousand frogs before you find a prince!”

- World $\mathcal{X}$ contains finite supply $S$ of princesses and $F$ of frogs.
- Each princess $\sigma$ has list $L(\sigma) \subseteq F$ of eligible frogs.
- Require injective map (no polyandry!)
  $f : S \rightarrow F$ with $f(\sigma) \in L(\sigma)$ for all princesses $\sigma$. 

Lemma 3.3 (Marriage Lemma)
A suitable injection $f : S \rightarrow F$ exists exactly when for all sets of princesses $A \subseteq S$

$$\#A \leq \#\cup\{L(\sigma) : \sigma \in A\}$$

(8)
Princesses and Frogs

“You gotta kiss a thousand frogs before you find a prince!”

- World $\mathcal{X}$ contains finite supply $S$ of princesses and $F$ of frogs.
- Each princess $\sigma$ has list $L(\sigma) \subseteq F$ of eligible frogs.
- Require injective map (no polyandry!)
  $f : S \rightarrow F$ with $f(\sigma) \in L(\sigma)$ for all princesses $\sigma$.

**Lemma 3.3 (Marriage Lemma)**

A suitable injection $f : S \rightarrow F$ exists exactly when for all sets of princesses $A \subseteq S$

$$\#A \leq \#L(A) \quad \text{for } L(A) = \bigcup \{ L(\sigma) : \sigma \in A \}.$$  

\hspace*{12cm} (8)
If Equation (8) fails then there are not enough frogs to go round some set $A$!
On the other hand Equation (8) suffices for just one princess ($\#S = 1$).
So use **induction** on $\#S \geq 1$:

- Princess $\sigma_1$ imperiously chooses first frog $f_1(\sigma_1)$ on her list $L(\sigma_1)$.
  Remaining princesses $S \setminus \{\sigma_1\}$ check frantically whether reduced lists $L(A) \setminus \{f_1(\sigma_1)\}$ will work (they use Equation (8) and induction).
  If not, (8) fails for some $\emptyset \neq A \subset S \setminus \{\sigma_1\}$, using the reduced lists: $\#L(A) \setminus \{f_1(\sigma_1)\} < \#A$ which forces $\#L(A) = \#A$ (use original Equation (8)).

- A set $A$ of aggrieved princesses splits off with their preferred frogs $L(A)$. The aggrieved set can solve their marriage problem (use Equation (8), induction).
If Equation (8) fails then there are not enough frogs to go round some set $A$!
On the other hand Equation (8) suffices for just one princess ($#S = 1$).
So use **induction** on $#S \geq 1$:

- Princess $\sigma_1$ imperiously chooses first frog $f_1(\sigma_1)$ on her list $L(\sigma_1)$. Remaining princesses $S \setminus \{\sigma_1\}$ check frantically whether reduced lists $L(\sigma) \setminus \{f_1(\sigma_1)\}$ will work (they use Equation (8) and **induction**).
If Equation (8) fails then there are not enough frogs to go round some set $A$!
On the other hand Equation (8) suffices for just one princess ($\#S = 1$).
So use **induction** on $\#S \geq 1$:

- Princess $\sigma_1$ imperiously chooses first frog $f_1(\sigma_1)$ on her list $L(\sigma_1)$. Remaining princesses $S \setminus \{\sigma_1\}$ check frantically whether reduced lists $L(\sigma) \setminus \{f_1(\sigma_1)\}$ will work (they use Equation (8) and **induction**).

- If not, (8) fails for some $\emptyset \neq A \subset S \setminus \{\sigma_1\}$, using the *reduced* lists:

$$\#L(A) \setminus \{f_1(\sigma_1)\} < \#A$$

which forces $\#L(A) = \#A$ (use original Equation (8)).
If Equation (8) fails then there are not enough frogs to go round some set $A$!
On the other hand Equation (8) suffices for just one princess ($\#S = 1$).
So use induction on $\#S \geq 1$:

- Princess $\sigma_1$ imperiously chooses first frog $f_1(\sigma_1)$ on her list $L(\sigma_1)$. Remaining princesses $S \setminus \{\sigma_1\}$ check frantically whether reduced lists $L(\sigma) \setminus \{f_1(\sigma_1)\}$ will work (they use Equation (8) and induction).

- If not, (8) fails for some $\emptyset \neq A \subset S \setminus \{\sigma_1\}$, using the reduced lists:

$$\#L(A) \setminus \{f_1(\sigma_1)\} < \#A$$

which forces $\#L(A) = \#A$ (use original Equation (8)).

- A set $A$ of aggrieved princesses splits off with their preferred frogs $L(A)$. The aggrieved set can solve their marriage problem (use Equation (8), induction).
What of remainder $S \setminus A$ with residual lists $\tilde{L}(\sigma) = L(\sigma) \setminus L(A)$? Observe:

$$\#(\tilde{L}(B) \cup L(A)) = \#\tilde{L}(B) + \#L(A)$$
$$= \#L(B \cup A)$$
$$\geq \#(B \cup A) = \#B + \#A = \#B + \#L(A).$$
What of remainder $S \setminus A$ with residual lists $	ilde{L}(\sigma) = L(\sigma) \setminus L(A)$? Observe:

$$\#(\tilde{L}(B) \cup L(A)) = \#\tilde{L}(B) + \#L(A)$$
$$= \#L(B \cup A)$$
$$\geq \#(B \cup A) = \#B + \#A = \#B + \#L(A).$$

So Equation (8) holds for remainder with residual lists: the proof can now be completed by induction.
What of remainder $S \setminus A$ with residual lists $\tilde{L}(\sigma) = L(\sigma) \setminus L(A)$? Observe:

$$\#(\tilde{L}(B) \cup L(A)) = \#\tilde{L}(B) + \#L(A)$$
$$= \#L(B \cup A)$$
$$\geq \#(B \cup A) = \#B + \#A = \#B + \#L(A).$$

So Equation (8) holds for remainder with residual lists: the proof can now be completed by induction.

“If you eat a live frog in the morning, nothing worse will happen to either of you for the rest of the day.”
Age brings avarice

“Kissing don’t last, cooking gold do”

The princesses, now happily married to their frogs, grow older. Inevitably they begin to exhibit mercenary tendencies — a sad but frequent tendency amongst royalty. Their major interest is now in gold.
Age brings avarice

“Kissing don’t last, cooking gold do”

The princesses, now happily married to their frogs, grow older. Inevitably they begin to exhibit mercenary tendencies — a sad but frequent tendency amongst royalty. Their major interest is now in gold.

- World $\mathcal{X}$ contains finite set $G$ of goldmines $g$ of capacity $\mu(g)$ tons per year respectively.
Age brings avarice

“Kissing don’t last, cooking gold do”

The princesses, now happily married to their frogs, grow older. Inevitably they begin to exhibit mercenary tendencies — a sad but frequent tendency amongst royalty. Their major interest is now in gold.

- World \( \mathcal{X} \) contains finite set \( G \) of goldmines \( g \) of capacity \( \mu(g) \) tons per year respectively.
- Each princess \( \sigma \in S \) requires \( \nu(\sigma) \) tons of gold per year from her list \( K(\sigma) \subseteq G \) of appropriate goldmines.
Age brings avarice

“Kissing don’t last, cooking gold do”

The princesses, now happily married to their frogs, grow older. Inevitably they begin to exhibit mercenary tendencies — a sad but frequent tendency amongst royalty. Their major interest is now in gold.

- World $\mathcal{X}$ contains finite set $G$ of goldmines $g$ of capacity $\mu(g)$ tons per year respectively.
- Each princess $\sigma \in S$ requires $\nu(\sigma)$ tons of gold per year from her list $K(\sigma) \subseteq G$ of appropriate goldmines.
- Require gold assignment using probability kernel

$$p : S \rightarrow \mathcal{P}(G),$$

$p(\sigma, \cdot)$ supported by $K(\sigma)$, with $\sum_\sigma \nu(\sigma)p(\sigma, g) \leq \mu(g)$. 
Lemma 3.4 (Marriage Lemma version 2)

A suitable probability kernel $p(\sigma, g)$ exists exactly when for all sets $A \subseteq S$ of princesses

$$\nu(A) \leq \mu(K(A)) \quad \text{for } K(A) = \bigcup \{K(\sigma) : \sigma \in A\}.$$  \hspace{1cm} (9)

Proof. Princesses aren't interested in small change and fractions of tons of gold. So we approximate by assuming $\mu(g), \nu(\sigma)$ are non-negative integers. (Can then proceed to rationals, then to reals!)

Now de-personalize those avaricious princesses: apply the Marriage Lemma to tons of gold instead of princesses!

"For the love of money is a root of all kinds of evil" — Paul of Tarsus, 1 Timothy 6:10 (NIV)
Lemma 3.4 (Marriage Lemma version 2)

A suitable probability kernel \( p(\sigma, g) \) exists exactly when for all sets \( A \subseteq S \) of princesses

\[
\nu(A) \leq \mu(K(A)) \quad \text{for } K(A) = \bigcup \{ K(\sigma) : \sigma \in A \}. \quad (9)
\]

Proof.
Princesses aren’t interested in small change and fractions of tons of gold. So we approximate by assuming \( \mu(g), \nu(\sigma) \) are non-negative integers. (Can then proceed to rationals, then to reals!)
Lemma 3.4 (Marriage Lemma version 2)

A suitable probability kernel \( p(\sigma, g) \) exists exactly when for all sets \( A \subseteq S \) of princesses

\[
\nu(A) \leq \mu(K(A)) \quad \text{for} \quad K(A) = \bigcup\{K(\sigma) : \sigma \in A\}.
\] (9)

Proof.
Princesses aren’t interested in small change and fractions of tons of gold. So we approximate by assuming \( \mu(g), \nu(\sigma) \) are non-negative integers. (Can then proceed to rationals, then to reals!)

Now de-personalize those avaricious princesses: apply the Marriage Lemma to tons of gold instead of princesses!
Lemma 3.4 (Marriage Lemma version 2)

A suitable probability kernel $p(\sigma, g)$ exists exactly when for all sets $A \subseteq S$ of princesses

$$\nu(A) \leq \mu(K(A)) \quad \text{for } K(A) = \bigcup\{K(\sigma) : \sigma \in A\}.$$  \hfill (9)

Proof.
Princesses aren’t interested in small change and fractions of tons of gold. So we approximate by assuming $\mu(g), \nu(\sigma)$ are non-negative integers. (Can then proceed to rationals, then to reals!)
Now de-personalize those avaricious princesses: apply the Marriage Lemma to tons of gold instead of princesses!

“For the love of money is a root of all kinds of evil”
— Paul of Tarsus, 1 Timothy 6:10 (NIV)
The princesses’ preferences for goldmines are geographically based. Suppose princess $\sigma$ is located at $x(\sigma) \in \mathcal{X}$, and goldmine $g$ at $y(g) \in \mathcal{X}$. Then

$$L(\sigma) = \{ g \in G : \text{dist}(x(\sigma), y(g)) < \varepsilon \}$$

for fixed $\varepsilon > 0$.

Avoid quarrels: assume $x(\sigma)$ are distinct and also $y(g)$ are distinct.
Capitalist geography

The princesses’ preferences for goldmines are geographically based. Suppose princess $\sigma$ is located at $x(\sigma) \in \mathcal{X}$, and goldmine $g$ at $y(g) \in \mathcal{X}$. Then

$$L(\sigma) = \{ g \in G : \text{dist}(x(\sigma), y(g)) < \varepsilon \} \text{ for fixed } \varepsilon > 0.$$

Avoid quarrels: assume $x(\sigma)$ are distinct and also $y(g)$ are distinct.

The king has been watching all these arrangements with interest. To make domestic life easier, he arranges for a convenient orbiting space-station to hold a modest amount of gold ($\varepsilon'$ tons, topped up each year). The princesses agree to add the space-station to their lists.
The second form of the Marriage Lemma shows, the royal demand for gold can be met so long as

\[ \sum_{\sigma: x(\sigma) \in B} \nu(\sigma) \leq \sum_{g: \text{dist}(g, \sigma) < \varepsilon \text{ for } x(\sigma) \in B} \mu(g) + \varepsilon'. \]

for any subset \( B \subseteq \mathcal{X} \).

\( nb: \ L(\{ \sigma : x(\sigma) \in B \}) = \{ g : \text{dist}(g, \sigma) < \varepsilon \text{ for } x(\sigma) \in B \}). \)
Back to probability

Replace princesses, gold demands, and locations by a $\mathcal{X}$-valued random variable $X$, distribution $P$. Similarly replace goldmines, annual yields, and locations by a probability distribution $Q$. When is close-coupling possible?

**Theorem 3.5 (Strassen’s Theorem)**

For probability distributions $P$, $Q$ on $\mathcal{X}$, we can find $\mathcal{X}$-valued random variables $X$, $Y$ with distributions $P$, $Q$ and such that

$$P[\text{dist}(X, Y) > \varepsilon] \leq \varepsilon'$$

exactly when, for all subsets $B \subseteq \mathcal{X}$,

$$P[B] \leq Q[x : \text{dist}(x, B) < \varepsilon] + \varepsilon'.$$
Proof.
Use the above to find suitable probability kernel $p(X, \cdot)$ and so construct suitable random variable $Y$ using
$L(Y|X) = p(X, \cdot)$,
Proof.
Use the above to find suitable probability kernel $p(X, \cdot)$ and so construct suitable random variable $Y$ using $\mathcal{L}(Y|X) = p(X, \cdot)$.

Strassen’s theorem says, we can couple $X$ and $Y$ to be close in a sense related to convergence in probability exactly when their distributions are close in a sense related to convergence in distribution, or weak convergence.
The above works for finite $\mathcal{X}$. For general $\mathcal{X}$ use measure-theory (Polish $\mathcal{X}$, measurable $\mathcal{B}$ ...).
Strassen’s theorem is useful when we want close-coupling (with just a small probability of being far away!). Suppose we want something different: a coupling with a positive chance of being exactly equal and otherwise no constraint. (Related to notion of *convergence stationnaire*, or “parking convergence”, for stochastic process theory.) Given two overlapping probability densities $f$ and $g$, we implement the coupling $(X, Y)$ as follows:

1. Compute $\alpha = \int (f \wedge g)(x) \, dx$, and with probability $\alpha$ return a draw of $X = Y$ from the density $(f \wedge g)/\alpha$.
2. Otherwise draw $X$ from $(f - f \wedge g)/(1 - \alpha)$ and $Y$ from $(g - f \wedge g)/(1 - \alpha)$.

This is related to the method of rejection sampling in stochastic simulation.

**EXERCISE 3.7**
Split chains and small sets II

From Doeblin’s time onwards, probabilists have applied this to study Markov chain transition probability kernels:

Definition 3.6 (Small set condition)
Let $X$ be a Markov chain on a state space $S$, transition kernel $p(x, dy)$. The set $C \subseteq S$ is a small set of order $k$ if for some probability measure $\nu$ and some $\alpha > 0$

\[
p^{(k)}(x, dy) \geq \mathbb{I}[C](x) \times \alpha \nu(dy).
\] (12)
It is a central result that small sets (possibly of arbitrarily high order) exist for any modestly regular Markov chain (proof in either of Nummelin 1984; Meyn and Tweedie 1993).

**Theorem 3.7**

*Let $X$ be a Markov chain on a non-discrete state space $S$, transition kernel $p(x, dy)$. Suppose an *(order 1)* small set $C \subseteq S$ exists. The dynamics of $X$ can be represented using a new Markov chain on $S \cup \{c\}$, for $c$ a regenerative pseudo-state.*
It is a central result that small sets (possibly of arbitrarily high order) exist for any modestly regular Markov chain (proof in either of Nummelin 1984; Meyn and Tweedie 1993).

**Theorem 3.7**

*Let $X$ be a Markov chain on a non-discrete state space $S$, transition kernel $p(x, dy)$. Suppose an (order 1) small set $C \subseteq S$ exists. The dynamics of $X$ can be represented using a new Markov chain on $S \cup \{c\}$, for $c$ a regenerative pseudo-state.*

For a more sophisticated theorem see Nummelin (1978), also Athreya and Ney (1978).
Small sets

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

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).

*Small sets of order 1* need not exist:
Small sets

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

*Small sets of order 1* need not exist:
Small sets

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).

Small sets of order 1 need not exist:

but will if
Small sets

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).

Small sets of order 1 need not exist:

but will if (a) the kernel \( p(x, dy) \) has a density…
Small sets

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).

*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.
Small sets and discretization of Markov chains

Theorem 3.8 (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

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

(13)
Small sets and discretization of Markov chains

Theorem 3.8 (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

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

(13)

So an evenly sampled Markov chain with transition density is a latent discrete Markov chain, with transition probabilities

\[
p_{ij} = \int g_i(u)f_j(u) \, du.
\]

(14)
Five is a sufficiently close approximation to infinity.
— Robert Firth

Approximation using coupling
  Skorokhod representation for weak convergence
  Central Limit Theorem and Brownian embedding
  Stein-Chen method for Poisson approximation
Skorokhod representation for weak convergence

We recall the inverse probability transform, and its use for simulation of random variables.

**Definition 4.1 (inverse probability transform)**

If

\[ F : \mathbb{R} \rightarrow [0, 1] \]

is increasing and right-continuous (ie: a distribution function) then we define its *inverse* as follows:

\[ F^{-1}(u) = \inf\{ t : F(t) \geq u \} \quad (15) \]

A graph of a distribution function helps to explain the reason for this definition!
Theorem 4.2
A real-valued random variable $X$ with distribution function

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

can be represented using the inverse probability transform

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

for $U$ a Uniform$[0, 1]$ random variable.
Theorem 4.2

A real-valued random variable $X$ with distribution function

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

can be represented using the inverse probability transform

\[ X = F^{-1}(U) , \]

for $U$ a Uniform$[0, 1]$ random variable.

Proof.
Consider

\[ \mathbb{P} \left[ F^{-1}(U) \leq x \right] = \mathbb{P} \left[ \inf \{ t : F(t) \geq U \} \leq x \right] \]
Theorem 4.2
A real-valued random variable $X$ with distribution function

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

can be represented using the inverse probability transform

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

for $U$ a Uniform$[0, 1]$ random variable.

Proof.
Consider

$$\mathbb{P} \left[ F^{-1}(U) \leq x \right] = \mathbb{P} \left[ \inf \{ t : F(t) \geq U \} \leq x \right] = \mathbb{P} [F(x) \geq U] - \ldots$$
Theorem 4.2

A real-valued random variable $X$ with distribution function

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

can be represented using the inverse probability transform

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

for $U$ a Uniform$[0, 1]$ random variable.

Proof.

Consider

$$\mathbb{P}\left[F^{-1}(U) \leq x\right] = \mathbb{P}\left[\inf\{t : F(t) \geq U\} \leq x\right] = \mathbb{P}[F(x) \geq U]$$

$$- \mathbb{P}[F(x) = U \text{ and } F(x - \varepsilon) = U \text{ for some } \varepsilon > 0].$$
Theorem 4.2

A real-valued random variable $X$ with distribution function

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

can be represented using the inverse probability transform

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

for $U$ a Uniform$[0,1]$ random variable.

Proof.

Consider

$$\mathbb{P}\left[F^{-1}(U) \leq x\right] = \mathbb{P}\left[\inf\{t : F(t) \geq U\} \leq x\right] = \mathbb{P}[F(x) \geq U]$$

$$- \mathbb{P}[F(x) = U \text{ and } F(x - \varepsilon) = U \text{ for some } \varepsilon > 0].$$
Theorem 4.2
A real-valued random variable $X$ with distribution function

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

can be represented using the inverse probability transform

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

for $U$ a Uniform$[0, 1]$ random variable.

Proof.
Consider

$$\mathbb{P}\left[F^{-1}(U) \leq x\right] = \mathbb{P}\left[\inf\{t : F(t) \geq U\} \leq x\right] = \mathbb{P}[F(x) \geq U]$$

$$= F(x).$$
If we do this with a single $U$ for an entire weakly convergent sequence of random variables, thus *coupling* the sequence, then we convert weak convergence to almost sure convergence.
If we do this with a single $U$ for an entire weakly convergent sequence of random variables, thus coupling the sequence, then we convert weak convergence to almost sure convergence.

For more general multivariate random variables (eg: with values in Polish spaces) we refer back to the methods of Strassen’s theorem.
Central Limit Theorem and Brownian embedding

Theorem 4.3
A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!
Theorem 4.3
A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!

Use the following basic facts about Brownian motion:
Central Limit Theorem and Brownian embedding

Theorem 4.3
A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!

Use the following basic facts about Brownian motion:

- Strong Markov property;
Central Limit Theorem and Brownian embedding

Theorem 4.3
A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!

Use the following basic facts about Brownian motion:

- Strong Markov property;
- Continuous sample paths;
Central Limit Theorem and Brownian embedding

Theorem 4.3
A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!

Use the following basic facts about Brownian motion:

- Strong Markov property;
- Continuous sample paths;
- Martingale property: $\mathbb{E} [B(S)] = 0$ if $S$ is a “bounded” stopping time ($B|_{[0,S]}$ bounded will do!);
Theorem 4.3

A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!

Use the following basic facts about Brownian motion:

- Strong Markov property;
- Continuous sample paths;
- Martingale property: $\mathbb{E} [B(S)] = 0$ if $S$ is a “bounded” stopping time ($B|_{[0,S]}$ bounded will do!);
- moreover, in that case $\mathbb{E} [B(S)^2] = \mathbb{E} [S]$;
Central Limit Theorem and Brownian embedding

Theorem 4.3
A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!

Use the following basic facts about Brownian motion:

- Strong Markov property;
- Continuous sample paths;
- Martingale property: $\mathbb{E}[B(S)] = 0$ if $S$ is a “bounded” stopping time ($B|_{[0,S]}$ bounded will do!);
- moreover, in that case $\mathbb{E}[B(S)^2] = \mathbb{E}[S]$;
- Scaling: $B(N\cdot)/\sqrt{N}$ has the same distribution as $B(\cdot)$.
Theorem 4.3
A zero-mean random variable $X$ of finite variance can be represented as $X = B(T)$ for a stopping time $T$ of finite mean. Thus Strong Law of Large Numbers and Brownian scaling imply CLT!

Use the following basic facts about Brownian motion:

- Strong Markov property;
- Continuous sample paths;
- Martingale property: $E[B(S)] = 0$ if $S$ is a “bounded” stopping time ($B|_{[0,S]}$ bounded will do!);
- moreover, in that case $E[B(S)^2] = E[S]$;
- Scaling: $B(N\cdot)/\sqrt{N}$ has the same distribution as $B(\cdot)$. 

EXERCISE 4.3
The promised proof of the CLT runs as follows:

\[ X_n = B(\sigma_1 + \ldots + \sigma_n) - B(\sigma_1 + \ldots + \sigma_{n-1}) \]
The promised proof of the CLT runs as follows:

\[ X_n = B(\sigma_1 + \ldots + \sigma_n) - B(\sigma_1 + \ldots + \sigma_{n-1}) \]

\[ \frac{\sigma_1 + \ldots + \sigma_N}{N \mathbb{E}[\sigma]} \to 1 \quad \text{almost surely by SLLN} \]
The promised proof of the CLT runs as follows:

\[ X_n = B(\sigma_1 + \ldots + \sigma_n) - B(\sigma_1 + \ldots + \sigma_{n-1}) \]

\[ \frac{\sigma_1 + \ldots + \sigma_N}{N \mathbb{E} \left[ \sigma \right]} \rightarrow 1 \quad \text{almost surely by SLLN} \]

(now convert construction to use a scaled path, and exploit continuity of the path)
The promised proof of the CLT runs as follows:

\[ X_n = B(\sigma_1 + \ldots + \sigma_n) - B(\sigma_1 + \ldots + \sigma_{n-1}) \]

\[ \frac{\sigma_1 + \ldots + \sigma_N}{N \mathbb{E}[\sigma]} \rightarrow 1 \quad \text{almost surely by SLLN} \]

(now convert construction to use a scaled path, and exploit continuity of the path)

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

and \( B(1) \) is Gaussian . . .
The promised proof of the CLT runs as follows:

\[
X_n = B(\sigma_1 + \ldots + \sigma_n) - B(\sigma_1 + \ldots + \sigma_{n-1})
\]

\[
\frac{\sigma_1 + \ldots + \sigma_N}{N \mathbb{E}[\sigma]} \to 1 \quad \text{almost surely by SLLN}
\]

(now convert construction to use a scaled path, and exploit continuity of the path)

\[
B\left(\frac{\sigma_1^{(N)} + \ldots + \sigma_N^{(N)}}{N \mathbb{E}[\sigma]}\right) \to B(1) \quad \text{almost surely}
\]

and \(B(1)\) is Gaussian . . .

\[
\text{LHS} \sim \frac{1}{\sqrt{N \mathbb{E}[\sigma]}} B(\sigma_1 + \ldots + \sigma_N) = \frac{1}{\sqrt{N \mathbb{E}[\sigma]}} \sum_1^N X_n.
\]
The promised proof of the CLT runs as follows:

\[ X_n = B(\sigma_1 + \ldots + \sigma_n) - B(\sigma_1 + \ldots + \sigma_{n-1}) \]

\[
\frac{\sigma_1 + \ldots + \sigma_N}{N \mathbb{E}[\sigma]} \to 1 \quad \text{almost surely by SLLN}
\]

(now convert construction to use a scaled path, and exploit continuity of the path)

\[ B\left(\frac{\sigma_1^{(N)} + \ldots + \sigma_N^{(N)}}{N \mathbb{E}[\sigma]}\right) \to B(1) \quad \text{almost surely} \]

and \( B(1) \) is Gaussian . . .

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

We deduce the following convergence in distribution:

\[ \frac{1}{\sqrt{N \mathbb{E}[\sigma]}} \sum_1^N X_n \to B(1). \]
This gives excellent intuition into extensions such as the Lindeberg condition: clearly the CLT should still work as long as $\sigma_1^{(N)} + \ldots + \sigma_n^{(N)}/N$ converges to a non-random constant (essentially a matter for SLLN for non-negative random variables!).
This gives excellent intuition into extensions such as the *Lindeberg condition*: clearly the CLT should still work as long as $\sigma_1^{(N)} + \ldots + \sigma_n^{(N)}/N$ converges to a non-random constant (essentially a matter for SLLN for non-negative random variables!).

We don’t even require independence: some kind of martingale conditions should (and do!) suffice.
This gives excellent intuition into extensions such as the *Lindeberg condition*: clearly the CLT should still work as long as \( \sigma_1^{(N)} + \ldots + \sigma_n^{(N)} / N \) converges to a non-random constant (essentially a matter for SLLN for non-negative random variables!).

We don’t even require independence: some kind of martingale conditions should (and do!) suffice.

But most significantly this approach suggests how to formulate a *Functional CLT*: the random-walk evolution \( X_1, X_2, \ldots \) should converge as random walk path to Brownian motion. An example of this kind of result is given in Kendall and Westcott (1987).
Stein-Chen method for Poisson approximation

Consider $W = \sum I_i$ (dependent binary $I_i$) thought to be approximated by a Poisson($\lambda$) random variable $\tilde{W}$. 

Then

$$\left| P[W \in A] - P[\tilde{W} \in A] \right| \leq \sup_n |g_A(n+1) - g_A(n)| \sum E[I_i] E[|U_i - V_i|].$$

Here $g_A(n)$ satisfies the Stein Estimating Equation

$$\lambda g_A(n+1) = ng_A(n) + I[n \in A] - P[\tilde{W} \in A].$$

(16)
Stein-Chen method for Poisson approximation

Consider $W = \sum I_i$ (dependent binary $I_i$) thought to be approximated by a Poisson($\lambda$) random variable $\tilde{W}$. Search for coupled $U_i, V_i$ such that $U_i$ has distribution of $W$, $V_i + 1$ has distribution of $W$ given $I_i = 1$. 

$$\left| P\left[ W \in A \right] - P\left[ \tilde{W} \in A \right] \right| \leq \sup_n |g_A(n+1) - g_A(n)| \sum E\left[ I_i \right] E\left[ |U_i - V_i| \right].$$

Here $g_A(n)$ satisfies the Stein Estimating Equation

$$\lambda g_A(n+1) = ng_A(n) + I(n \in A) - P\left[ \tilde{W} \in A \right].$$

(16)
Stein-Chen method for Poisson approximation

Consider \( W = \sum I_i \) (dependent binary \( I_i \)) thought to be approximated by a Poisson(\( \lambda \)) random variable \( \tilde{W} \).

Search for coupled \( U_i, V_i \) such that \( U_i \) has distribution of \( W \), \( V_i + 1 \) has distribution of \( W \) given \( I_i = 1 \). Then

\[
\left| \mathbb{P} [W \in A] - \mathbb{P} [\tilde{W} \in A] \right| \leq \sup_n \left| g_A(n + 1) - g_A(n) \right| \sum \mathbb{E} [I_i] \mathbb{E} [\|U_i - V_i\|].
\]
Stein-Chen method for Poisson approximation

Consider $W = \sum I_i$ (dependent binary $I_i$) thought to be approximated by a Poisson($\lambda$) random variable $\tilde{W}$. Search for coupled $U_i, V_i$ such that $U_i$ has distribution of $W$, $V_i + 1$ has distribution of $W$ given $I_i = 1$. Then

\[
\left| \Pr[W \in A] - \Pr[\tilde{W} \in A] \right| \leq \sup_n \left| g_A(n + 1) - g_A(n) \right| \sum \mathbb{E}[I_i] \mathbb{E}[|U_i - V_i|].
\]

Here $g_A(n)$ satisfies the Stein Estimating Equation
Stein-Chen method for Poisson approximation

Consider $W = \sum I_i$ (dependent binary $I_i$) thought to be approximated by a Poisson($\lambda$) random variable $\tilde{W}$.

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

$$\left| \mathbb{P} [W \in A] - \mathbb{P} [\tilde{W} \in A] \right| \leq \sup_n |g_A(n + 1) - g_A(n)| \sum \mathbb{E} [I_i] \mathbb{E} [|U_i - V_i|].$$

Here $g_A(n)$ satisfies the Stein Estimating Equation

$$\lambda g_A(n + 1) = ng_A(n) + \mathbb{I} [n \in A] - \mathbb{P} [\tilde{W} \in A]. \quad (16)$$
Consider

\[ \left| P[W \in A] - P[\tilde{W} \in A] \right| \leq \sup_n |g_A(n + 1) - g_A(n)| \sum E[l_i] E|U_i - V_i| . \]
Consider

$$\left| \Pr [W \in A] - \Pr [\tilde{W} \in A] \right| \leq \sup_n \left| g_A(n+1) - g_A(n) \right| \sum \mathbb{E} [l_i] \mathbb{E} [|U_i - V_i|].$$

After analysis (see Barbour, Holst, and Janson 1992)

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

(17)

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

(18)
Consider

\[
\left| P[W \in A] - P[\tilde{W} \in A] \right| \leq 
\sup_n |g_A(n + 1) - g_A(n)| \sum E[I_i] E[|U_i - V_i|].
\]

After analysis (see Barbour, Holst, and Janson 1992)

\[
\sup_n |g_A(n)| \leq \min \left\{ 1, \frac{1}{\sqrt{\lambda}} \right\} \quad (17)
\]

\[
\sup_n |g_A(n + 1) - g_A(n)| \leq \frac{1 - e^{-\lambda}}{\lambda} \leq \min \left\{ 1, \frac{1}{\lambda} \right\} \quad (18)
\]

Even better, if \( U_i \geq V_i \) (say), the sum collapses:

\[
\sum E[I_i] E[|U_i - V_i|] = \lambda - \text{Var}[W]. \quad (19)
\]
“There are only 10 types of people in this world: those who understand binary and those who don’t.”
Lecture 5: Mixing of Markov chains

What is (15 minus three times five) plus
Lecture 5: Mixing of Markov chains

What is (15 minus three times five) plus

(20 minus four times five) plus

(36 minus nine times four) plus

(72 minus nine times eight) plus

(98 minus eight times twelve) plus

(54 minus seven times eight)?
Lecture 5: Mixing of Markov chains

What is (15 minus three times five) plus

- (20 minus four times five) plus
- (36 minus nine times four) plus
Lecture 5: Mixing of Markov chains

What is (15 minus three times five) plus
► (20 minus four times five) plus
► (36 minus nine times four) plus
► (72 minus nine times eight) plus

A lot of work for nothing.
Lecture 5: Mixing of Markov chains

What is (15 minus three times five) plus
► (20 minus four times five) plus
► (36 minus nine times four) plus
► (72 minus nine times eight) plus
► (98 minus eight times twelve) plus
Lecture 5: Mixing of Markov chains

What is \((15 \text{ minus three times five}) \text{ plus} \) 
\((20 \text{ minus four times five}) \text{ plus} \) 
\((36 \text{ minus nine times four}) \text{ plus} \) 
\((72 \text{ minus nine times eight}) \text{ plus} \) 
\((98 \text{ minus eight times twelve}) \text{ plus} \) 
\((54 \text{ minus seven times eight})\)?
Lecture 5: Mixing of Markov chains

What is (15 minus three times five) plus
► (20 minus four times five) plus
► (36 minus nine times four) plus
► (72 minus nine times eight) plus
► (98 minus eight times twelve) plus
► (54 minus seven times eight)?

A lot of work for nothing.

http://cnonline.net/~TheCookieJar/
What is (15 minus three times five) plus
▶ (20 minus four times five) plus
▶ (36 minus nine times four) plus
▶ (72 minus nine times eight) plus
▶ (98 minus eight times twelve) plus
▶ (54 minus seven times eight)?

A lot of work for nothing.

http://cnonline.net/~TheCookieJar/

Mixing of Markov chains
  Coupling inequality
  Very simple mixing example
  Slice sampler
  Strong stationary times
Coupling inequality

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

$$\text{dist}_{tv}(\mathcal{L}(X_n), \pi) \leq \max_y \{ \mathbb{P}[T_{x,y} > t] \}. \quad (20)$$
Suppose $X$ is a Markov chain, with equilibrium distribution $\pi$, for which we can produce a coupling between any two points $x$, $y$, which succeeds at time $T_{x,y} < \infty$. Then

$$\text{dist}_{tv}(\mathcal{L}(X_n), \pi) \leq \max_y \{ \mathbb{P} [T_{x,y} > t] \}.$$  \hspace{1cm} (20)

Useful bounds depend on finding and analyzing the right coupling!
Coupling inequality

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

$$\text{dist}_{tv}(\mathcal{L}(X_n), \pi) \leq \max_y \{P[T_{x,y} > t]\}.$$  \hspace{1cm} (20)

Useful bounds depend on finding and analyzing the right coupling!

Coupling inequality

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

$$\text{dist}_{tv}(\mathcal{L}(X_n), \pi) \leq \max_y \{\mathbb{P}[T_{x,y} > t]\}.$$  \hspace{1cm} (20)

Useful bounds depend on finding and analyzing the right coupling!


The coupling inequality is also the basis for an empirical approach to convergence estimation (Johnson 1996).
Very simple mixing example

Utterly the simplest case: a Markov chain on \{0, 1\}.

Let \(X, Y\) start at 0, 1, with transitions
- 0 \(\rightarrow\) 1 at rate \(1/\alpha\),
- and 1 \(\rightarrow\) 0 at rate \(1/\alpha\).
Very simple mixing example

Utterly the simplest case: a Markov chain on \{0, 1\}.

Let \(X, Y\) start at 0, 1, with transitions

- \(0 \to 1\) at rate \(1/\alpha\),
- and \(1 \to 0\) at rate \(1/\alpha\).

Supply (a) Poisson process (rate \(1/\alpha\)) of \(0 \to 1\) transitions,
(b) *ditto* of \(1 \to 0\) transitions. Apply them where applicable to \(X, Y\). Clearly \(X, Y\) have desired distributions.
Very simple mixing example

Utterly the simplest case: a Markov chain on \{0, 1\}.

Let \( X, Y \) start at 0, 1, with transitions
- 0 \( \rightarrow \) 1 at rate \( 1/\alpha \),
- and 1 \( \rightarrow \) 0 at rate \( 1/\alpha \).

Supply (a) Poisson process (rate \( 1/\alpha \)) of 0 \( \rightarrow \) 1 transitions, (b) ditto of 1 \( \rightarrow \) 0 transitions. Apply them where applicable to \( X, Y \). Clearly \( X, Y \) have desired distributions.

**Coupling** happens at first instant of combined Poisson process, hence at Exponential\((2/\alpha)\) random time.
Very simple mixing example

Utterly the simplest case: a Markov chain on \{0, 1\}.

Let \( X, Y \) start at 0, 1, with transitions

- 0 → 1 at rate \( \frac{1}{\alpha} \),
- and 1 → 0 at rate \( \frac{1}{\alpha} \).

Supply (a) Poisson process (rate \( \frac{1}{\alpha} \)) of 0 → 1 transitions, (b) ditto of 1 → 0 transitions. Apply them where applicable to \( X, Y \). Clearly \( X, Y \) have desired distributions.

**Coupling** happens at first instant of combined Poisson process, hence at Exponential\( (\frac{2}{\alpha}) \) random time.

The coupled equilibrium chain begins in between \( X \) and \( Y \), and is equal to both at time of coupling. Hence we can estimate rate of convergence to equilibrium . . .
Slice sampler

Task is to draw from density $f(x)$. Here is a one-dimensional example! (But note, this is only interesting because it can be made to work in many dimensions . . . ) Suppose $f$ unimodal. Define $g_0(y), g_1(y)$ implicitly by: $[g_0(y), g_1(y)]$ is the superlevel set $\{x : f(x) \geq y\}$. Alternate between drawing $y$ uniformly from $[0, f(x)]$ and drawing $x$ uniformly from $[g_0(y), g_1(y)]$.

Roberts and Rosenthal (1999, Theorem 12) show rapid convergence (order of 530 iterations!) under a specific variation of log-concavity.
Definition 5.1 (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 $\pi$) if

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

Application to shuffling pack of \( n \) cards by random transpositions (Broder): based on notion of checked cards.
Strong stationary times II

Application to shuffling pack of $n$ cards by random transpositions (*Broder*): based on notion of checked cards.

Transpose by choosing 2 cards at random; LH, RH.

Inductive claim: given number of checked cards, positions in pack of checked cards, list of values of cards, then the map of checked card to value is uniformly random.

Let $T_m$ be time when $m$ cards checked. Then (induction)

$$T_m = \sum_{n - 1}^{m} T_m$$

is a strong stationary time.

EXERCISE 5.4
Strong stationary times II

Application to shuffling pack of $n$ cards by random transpositions (Broder): based on notion of checked cards.

Transpose by choosing 2 cards at random; LH, RH. Check card pointed to by LH if

$\text{EXERCISE 5.4}$
Strong stationary times II

Application to shuffling pack of $n$ cards by random transpositions (*Broder*): based on 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;

Inductive claim: given number of checked cards, positions in pack of checked cards, list of values of cards, then the map of checked card to value is uniformly random.

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

EXERCISE 5.4
Strong stationary times II

Application to shuffling pack of $n$ cards by random transpositions (Broder): based on 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, then the map of checked card to value is uniformly random.

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

EXERCISE 5.4
Application to shuffling pack of $n$ cards by random transpositions \((\textit{Broder})\): based on 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, \emph{list} of values of cards, then the map of checked card to value is uniformly random.
Strong stationary times II

Application to shuffling pack of \( n \) cards by random transpositions (Broder): based on 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, then 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.
Strong stationary times III

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

Exercise 5.5
Strong stationary times III

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

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

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.

EXERCISE 5.5
Strong stationary times III

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

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

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 (or a more careful probabilistic construction) the correct asymptotic is $\frac{1}{2} n \log n$. 

EXERCISE 5.5
They told me you had proven it
About a month before.
The proof was valid, more or less
But rather less than more.
They told me you had proven it
   About a month before.
The proof was valid, more or less
   But rather less than more.

He sent them word that we would try
   To pass where they had failed
And after we were done, to them
   The new proof would be mailed.
They told me you had proven it
   About a month before.
The proof was valid, more or less
   But rather less than more.

He sent them word that we would try
   To pass where they had failed
And after we were done, to them
   The new proof would be mailed.

My notion was to start again
   Ignoring all they’d done
We quickly turned it into code
   To see if it would run.
They told me you had proven it
   About a month before.
The proof was valid, more or less
   But rather less than more.

He sent them word that we would try
   To pass where they had failed
And after we were done, to them
   The new proof would be mailed.

My notion was to start again
   Ignoring all they’d done
We quickly turned it into code
   To see if it would run.

When they discovered our results
   Their blood began to freeze
Instead of understanding it
   We’d run thirteen MCMC’s.
They told me you had proven it
   About a month before.
The proof was valid, more or less
   But rather less than more.

He sent them word that we would try
   To pass where they had failed
And after we were done, to them
   The new proof would be mailed.

My notion was to start again
   Ignoring all they’d done
We quickly turned it into code
   To see if it would run.

When they discovered our results
   Their blood began to freeze
Instead of understanding it
   We’d run thirteen MCMC’s.

Don’t tell a soul about all this
   For it must ever be
A secret, kept from all the rest
   Between yourself and me.
Lecture 6: The Coupling Zoo

There is a whole variety of different ways of constructing couplings. Here is a zoo containing some representative specimens.

But my Totem saw the shame; from his ridgepole shrine he came,  
And he told me in a vision of the night:  
“There are nine and sixty ways of constructing tribal lays,  
And every single one of them is right!”

_In the Neolithic Age — Rudyard Kipling_

The Coupling Zoo  
Zoo taxonomy
### The Coupling Zoo

A politically inspired taxonomy

<table>
<thead>
<tr>
<th>Coupling method</th>
<th>Before coupling happens …</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent (Doeblin)</td>
<td>Be independent.</td>
</tr>
</tbody>
</table>

The easiest coupling method to study!
The Coupling Zoo
A politically inspired taxonomy

**Coupling method:**
- Independent (Doeblin)
- Synchronous
- Reflection

**Before coupling happens ...**
- Be independent.
- Do the **same**.
- Do the **opposite**.

---

**Issue:** *how* to synchronize or reflect Markov chain jumps?

*Without* linear structure in the state-space, the right definition is not obvious ...
The Coupling Zoo
A politically inspired taxonomy

**Coupling method:**
- Independent (Doeblin)
- Synchronous
- Reflection
- Ornstein (random walks)

**Before coupling happens . . .**
- Be independent.
- Do the same.
- Do the opposite.
- Be a liberal democrat: agree big steps, otherwise be independent . . .

Great advantage: converts random walks to random walks with symmetric bounded increments!
### The Coupling Zoo
A politically inspired taxonomy

<table>
<thead>
<tr>
<th>Coupling method:</th>
<th>Before coupling happens . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent (Doeblin)</td>
<td>Be independent.</td>
</tr>
<tr>
<td>Synchronous</td>
<td>Do the same.</td>
</tr>
<tr>
<td>Reflection</td>
<td>Do the opposite.</td>
</tr>
<tr>
<td>Ornstein (random walks)</td>
<td>Be a liberal democrat: agree big steps, otherwise be independent . .</td>
</tr>
<tr>
<td>Mineka (random walks)</td>
<td>. . . be new labour: difference is simple symmetric random walk.</td>
</tr>
</tbody>
</table>
The Coupling Zoo
A politically inspired taxonomy

**Coupling method:**
- Independent (Doeblin)
- Synchronous
- Reflection
- Ornstein (random walks)
- Mineka (random walks)
- Vasershtein

**Before coupling happens . . .**
- Be independent.
- Do the same.
- Do the opposite.
- Be a liberal democrat: agree big steps, otherwise be independent . . .
- . . . be new labour: difference is simple symmetric random walk.
- Maximize immediate success.

The Vasershtein coupling is related to small-set coupling:
look at the overlap between jump distributions.
## The Coupling Zoo

A politically inspired taxonomy

<table>
<thead>
<tr>
<th>Coupling method:</th>
<th>Before coupling happens ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent (Doeblin)</td>
<td>Be independent.</td>
</tr>
<tr>
<td>Synchronous</td>
<td>Do the same.</td>
</tr>
<tr>
<td>Reflection</td>
<td>Do the opposite.</td>
</tr>
<tr>
<td>Ornstein (random walks)</td>
<td>Be a liberal democrat: agree big steps, otherwise be independent ...</td>
</tr>
<tr>
<td>Mineka (random walks)</td>
<td>... be new labour: difference is simple symmetric random walk.</td>
</tr>
<tr>
<td>Vasershtein</td>
<td>Maximize immediate success.</td>
</tr>
<tr>
<td>Time-changed</td>
<td>Run chains at varying rates.</td>
</tr>
</tbody>
</table>

Vary time-rates to make sure coupling happens before either of the processes hits a fixed boundary (Kendall 1994).
## The Coupling Zoo
* A politically inspired taxonomy

### Coupling method:
- Independent (Doeblin)
- Synchronous
- Reflection
- Ornstein (random walks)
- Mineka (random walks)
- Vasershtein
- Time-changed
- **Maximal (non-adaptive)**

### Before coupling happens . . .
- Be independent.
- Do the same.
- Do the opposite.
- Be a liberal democrat: agree big steps, otherwise be independent . . .
  . . . be new labour: difference is simple symmetric random walk.
- Maximize immediate success.
- Run chains at varying rates.
- Practice **inside-trading**.
  (Griffeath 1975; Goldstein 1979)

---

Suppose we allow non-adaptive coupling strategies (one process depends on *future* of other!). This relates to bounded *space-time harmonic functions.*
# The Coupling Zoo

A politically inspired taxonomy

<table>
<thead>
<tr>
<th>Coupling method:</th>
<th>Before coupling happens . . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent (Doeblin)</td>
<td>Be independent.</td>
</tr>
<tr>
<td>Synchronous</td>
<td>Do the same.</td>
</tr>
<tr>
<td>Reflection</td>
<td>Do the opposite.</td>
</tr>
<tr>
<td>Ornstein (random walks)</td>
<td>Be a liberal democrat: agree big steps, otherwise be independent . . .</td>
</tr>
<tr>
<td>Mineka (random walks)</td>
<td>. . . be new labour: difference is simple symmetric random walk.</td>
</tr>
<tr>
<td>Vasershtein</td>
<td>Maximize immediate success.</td>
</tr>
<tr>
<td>Time-changed</td>
<td>Run chains at varying rates.</td>
</tr>
<tr>
<td>Maximal (non-adaptive)</td>
<td>Practice inside-trading.</td>
</tr>
</tbody>
</table>

**Shift**

(Griffeath 1975; Goldstein 1979)

Do the **same** thing at **different times**!

(Aldous and Thorisson 1993)

Shift-coupling requires simply that processes $X, Y$ should agree after **different** stopping times $S, T$. Related to bounded **harmonic functions**.
## The Coupling Zoo
A politically inspired taxonomy

<table>
<thead>
<tr>
<th>Coupling method:</th>
<th>Before coupling happens …</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent (Doeblin)</td>
<td>Be independent.</td>
</tr>
<tr>
<td>Synchronous</td>
<td>Do the same.</td>
</tr>
<tr>
<td>Reflection</td>
<td>Do the opposite.</td>
</tr>
<tr>
<td>Ornstein (random walks)</td>
<td>Be a liberal democrat: agree big steps, otherwise be independent …</td>
</tr>
<tr>
<td>Mineka (random walks)</td>
<td>… be new labour: difference is simple symmetric random walk.</td>
</tr>
<tr>
<td>Vasershtein</td>
<td>Maximize immediate success.</td>
</tr>
<tr>
<td>Time-changed</td>
<td>Run chains at varying rates.</td>
</tr>
<tr>
<td>Maximal (non-adaptive)</td>
<td>Practice inside-trading.</td>
</tr>
<tr>
<td>Shift</td>
<td>Do the same thing at different times!</td>
</tr>
</tbody>
</table>

(Griffeath 1975; Goldstein 1979)

(Aldous and Thorisson 1993)
So which coupling is the best?
It all depends what you want to do . . .

What did the Colonel’s Lady think?
Nobody never knew.
Somebody asked the Sergeant’s Wife,
An’ she told ’em true!
When you get to a man in the case,
They’re like as a row of pins –
For the Colonel’s Lady an’ Judy O’Grady
Are sisters under their skins!

The Ladies — Rudyard Kipling, Barrack Room Ballads
“You can arrive (mayan arivan on-when) for any sitting you like without prior (late fore-when) reservation because you can book retrospectively, as it were when you return to your own time. (you can have on-book haventa forewhen presooning returningwenta retrohome.)”

Douglas Adams — Hitchhiker’s Guide to the Galaxy
The constructive nature of probabilistic coupling ("to build $Y$ using the randomness of $X$") makes it close in spirit to the task of constructing good stochastic simulations. Recently the link between coupling and simulation has been strengthened in striking ways, resulting in so-called "exact" or "perfect simulation".

Häggström (2002) includes discussion of some of these ideas at the level of a student monograph.

See also

- Aldous and Fill (200x);
- Møller and Waagepetersen (2003);

and

http://research.microsoft.com/~dbwilson/exact/
More on mixing

Recall (continuous-time!) random walk on \{0, 1\}.

**Construction:** Supply (a) Poisson process (rate \(1/\alpha\)) of 0 \(\rightarrow\) 1 transitions, (b) *ditto* of 1 \(\rightarrow\) 0 transitions. Apply them where applicable to \(X\) started at 0, \(Y\) started at 1.
More on mixing

Recall (continuous-time!) random walk on \{0, 1\}.

**Construction:** Supply (a) Poisson process (rate \(1/\alpha\)) of 0 \(\rightarrow\) 1 transitions, (b) *ditto* of 1 \(\rightarrow\) 0 transitions. Apply them where applicable to \(X\) started at 0, \(Y\) started at 1. Clearly \(X, Y\) have desired distributions.
More on mixing

Recall (continuous-time!) random walk on \{0, 1\}.

Construction: Supply (a) Poisson process (rate $1/\alpha$) of $0 \to 1$ transitions, (b) *ditto* of $1 \to 0$ transitions. Apply them where applicable to $X$ started at 0, $Y$ started at 1.

Clearly $X$, $Y$ have desired distributions.

Questions:
Does it make sense to return first coupled value? (Yes here, no in “nearly every” other case.)
More on mixing

Recall (continuous-time!) random walk on \( \{0, 1\} \).

**Construction:** Supply (a) Poisson process (rate \( 1/\alpha \)) of \( 0 \to 1 \) transitions, (b) *ditto* of \( 1 \to 0 \) transitions. Apply them where applicable to \( X \) started at 0, \( Y \) started at 1.

Clearly \( X, Y \) have desired distributions.

**Questions:**

Does it make sense to return first coupled value? (Yes here, no in “nearly every” other case.)

But suppose I run algorithm from \([-T, 0]\) for increasing \( T \), instead of from \([0, T]\) for increasing \( T \)? (This *will* work always: this is **CFTP**.)
More on mixing

Recall (continuous-time!) random walk on \{0, 1\}.

**Construction:** Supply (a) Poisson process (rate \(1/\alpha\)) of 0 \(\rightarrow\) 1 transitions, (b) *ditto* of 1 \(\rightarrow\) 0 transitions. Apply them where applicable to \(X\) started at 0, \(Y\) started at 1.

Clearly \(X, Y\) have desired distributions.

**Questions:**

Does it make sense to return first coupled value? (Yes here, no in “nearly every” other case.)

But suppose I run algorithm from \([-T, 0]\) for increasing \(T\), instead of from \([0, T]\) for increasing \(T\)? (This *will* work always: this is \(CFTP\).)
Conventionally, the Markov chain is extended into the future $T$ using innovations.
Conventionally, the Markov chain is extended into the future $T$ using innovations.

Equilibrium in the unreachably far infinite future $T \to \infty$. 
Conventionally, the Markov chain is extended into the future $T$ using innovations.

Equilibrium in the unreachably far infinite future $T \to \infty$.

Ergodic theorists have always known of an alternative:
Classic CFTP

Conventionally, the Markov chain is extended into the future $T$ using innovations. Equilibrium in the unreachably far infinite future $T \to \infty$. Ergodic theorists have always known of an alternative:

view the chain as produced at time $t = 0$ by innovations acting from some starting point $-T$ in the past:
Conventionally, the Markov chain is extended into the future $T$ using innovations. Equilibrium in the unreachably far infinite future $T \to \infty$.

Ergodic theorists have always known of an alternative:

- view the chain as produced at time $t = 0$ by innovations acting from some starting point $-T$ in the past:
  - flows not processes, dependence on initial conditions, $X_0$ at 0 no longer Markov as $-T \to -\infty$, **BUT** . . .
Conventionally, the Markov chain is extended into the future $T$ using innovations. Equilibrium in the unreachable far infinite future $T \to \infty$. Ergodic theorists have always known of an alternative:

view the chain as produced at time $t = 0$ by innovations acting from some starting point $-T$ in the past:

flows not processes, dependence on initial conditions, $X_0$ at 0 no longer Markov as $-T \to -\infty$, but:

if the value of $X_0$ stabilizes to a single value then this value is a draw from equilibrium!
Theorem 7.1
If coalescence is almost sure then CFTP delivers a sample from the equilibrium distribution of the Markov chain \(X\) corresponding to the random input-output maps \(F_{(-u,v]}\).
Theorem 7.1
If coalescence is almost sure then CFTP delivers a sample from the equilibrium distribution of the Markov chain $X$ corresponding to the random input-output maps $F_{(-u,v]}$.

Proof.
For each $[-n, \infty)$ use input-output maps $F_{(-n,t]}$. Assume finite coalescence time $-n = -T$ for $F_{(-n,0]}$. 

Theorem 7.1
If coalescence is almost sure then CFTP delivers a sample from the equilibrium distribution of the Markov chain $X$ corresponding to the random input-output maps $F_{(-u,v]}$.

Proof.
For each $[-n, \infty)$ use input-output maps $F_{(-n,t]}$. Assume finite coalescence time $-n = -T$ for $F_{(-n,0]}$. Then

$$X_{t}^{-n} = F_{(-n,t]}(0) \quad \text{for} \quad -n \leq t.$$  
$$X_{0}^{-n} = X_{0}^{-T} \quad \text{whenever} \quad -n \leq -T;$$  
$$\mathcal{L}(X_{0}^{-n}) = \mathcal{L}(X_{n}^{0}).$$
Theorem 7.1
If coalescence is almost sure then CFTP delivers a sample from the equilibrium distribution of the Markov chain $X$ corresponding to the random input-output maps $F_{(-u,v]}$.

Proof.
For each $[-n, \infty)$ use input-output maps $F_{(-n,t]}$. Assume finite coalescence time $-n = -T$ for $F_{(-n,0]}$. Then

$$X_{t}^{-n} = F_{(-n,t]}(0) \quad \text{for } -n \leq t.$$  
$$X_{0}^{-n} = X_{0}^{-T} \quad \text{whenever } -n \leq -T;$$  
$$\mathcal{L}(X_{0}^{-n}) = \mathcal{L}(X_{0}^{0}).$$

If $X$ converges to an equilibrium $\pi$ then

$$\text{dist}_{tv}(\mathcal{L}(X_{0}^{-T}), \pi) = \lim_{n} \text{dist}_{tv}(\mathcal{L}(X_{0}^{-n}), \pi) = \lim_{n} \text{dist}_{tv}(\mathcal{L}(X_{0}^{0}), \pi) = 0.$$
Theorem 7.1
If coalescence is almost sure then CFTP delivers a sample from the equilibrium distribution of the Markov chain $X$ corresponding to the random input-output maps $F_{(-u,v]}$.

Proof.
For each $[-n,\infty)$ use input-output maps $F_{(-n,t]}$. Assume finite coalescence time $-n = -T$ for $F_{(-n,0]}$. Then

$$X_{t}^{-n} = F_{(-n,t]}(0) \quad \text{for} \quad -n \leq t.$$  
$$X_{0}^{-n} = X_{0}^{-T} \quad \text{whenever} \quad -n \leq -T;$$  
$$\mathcal{L}(X_{0}^{-n}) = \mathcal{L}(X_{n}^{0}).$$

If $X$ converges to an equilibrium $\pi$ then

$$\text{dist}_{tv}(\mathcal{L}(X_{0}^{-T}), \pi) = \lim_{n} \text{dist}_{tv}(\mathcal{L}(X_{0}^{-n}), \pi) = \lim_{n} \text{dist}_{tv}(\mathcal{L}(X_{n}^{0}), \pi) = 0.$$
Falling leaves

Kendall and Thönnes (1999) describe a visual and geometric application of CFTP in mathematical geology: this particular example being well-known to workers in the field previous to the introduction of CFTP itself.
Falling leaves

Kendall and Thönnes (1999) describe a visual and geometric application of $CFTP$ in mathematical geology: this particular example being well-known to workers in the field previous to the introduction of $CFTP$ itself.

Occlusion $CFTP$ for the falling leaves of Fontainbleau.
Falling leaves

Kendall and Thönnes (1999) describe a visual and geometric application of CFTP in mathematical geology: this particular example being well-known to workers in the field previous to the introduction of CFTP itself.

Occlusion CFTP for the falling leaves of Fontainbleau.

(Why “occlusion”? David Wilson’s terminology: this CFTP algorithm builds up the result piece-by-piece with no back-tracking.)
The original Propp and Wilson (1996) idea showed how to make exact draws from the critical Ising model. A rather simpler application uses the heat-bath sampler to make exact draws from the sub-critical Ising model.
The original Propp and Wilson (1996) idea showed how to make exact draws from the critical Ising model. A rather simpler application uses the heat-bath sampler to make exact draws from the sub-critical Ising model.

Classic CFTP for the Ising model (simple, sub-critical case). Heat-bath dynamics run from past; compare results from maximal and minimal starting conditions.
Approaches based on Swendsen-Wang ideas work for critical case (Huber 2003).
Ising model (II)

Approaches based on Swendsen-Wang ideas work for critical case (Huber 2003).

On the other hand, conditioning by noisy data conveniently reduces the difficulties caused by phase-transition phenomena.
Ising model (II)

Approaches based on Swendsen-Wang ideas work for critical case (Huber 2003).

On the other hand, conditioning by noisy data conveniently reduces the difficulties caused by phase-transition phenomena.

Classic CFTP for the Ising model conditioned by noisy data. Without influence from data (“external magnetic field”) this Ising model would be super-critical.
Small-set CFTP

A common misconception is that CFTP requires an ordered state-space. Green and Murdoch (1999) showed how to use small sets to carry out CFTP when the state-space is continuous with no helpful ordering. Their prescription includes the use of a partition by several small sets, to speed up coalescence.

Small set CFTP in nearly the simplest possible case: a triangular kernel over [0, 1].
“What’s past is –”
“– is very much in the present. Brigadier, you never
did understand the interrelation of time.”
— The Brigadier and the Doctor, in “Mawdryn Undead”
Lecture 8: Perfection (CFTP II)

- “I think there may be one or two steps in your logic that I have failed to grasp, Mister Stibbons,” said the Archchancellor coldly. “I suppose you’re not intending to shoot your own grandfather, by any chance?”
- “Of course not!” snapped Ponder, “I don’t even know what he looked like. He died before I was born.”
- “Ah-hah!”

The faculty members of Unseen University discuss time travel

— The Last Continent, Terry Pratchett

Perfection (CFTP II)
Point patterns
Dominated CFTP for area-interaction point processes
Fast attractive area-interaction CFTP
Layered Multishift CFTP
Perpetuities
Dominated CFTP

Successful application of CFTP usually requires the combination of a number of tricks. We have already seen examples in which monotonicity plays a part, either openly (random walk, Ising) or concealed (dead leaves).
Dominated CFTP

Successful application of \textit{CFTP} usually requires the combination of a number of tricks. We have already seen examples in which monotonicity plays a part, either openly (random walk, Ising) or concealed (dead leaves).

One of the major issues concerns how to deal with Markov chains that are not “bounded” (technically speaking, not uniformly ergodic, in sense of Defn. 10.2).\(^2\)

\(^2\)“uniformly ergodic”: rate of convergence to equilibrium doesn’t depend on starting point.
Dominated CFTP

Successful application of CFTP usually requires the combination of a number of tricks. We have already seen examples in which monotonicity plays a part, either openly (random walk, Ising) or concealed (dead leaves).

One of the major issues concerns how to deal with Markov chains that are not “bounded” (technically speaking, not uniformly ergodic, in sense of Defn. 10.2).\(^2\)

Consider the point-pattern-valued birth-death processes often used in modelling spatial point processes . . .

\(^2\)“uniformly ergodic”: rate of convergence to equilibrium doesn’t depend on starting point.
Point processes

Algorithm to generate a random point pattern \( X \) (see, eg, Stoyan et al. 1995). Repeat the following:

- suppose the current value of \( X \) is \( X = x \):
  
  - after Exponential \( \lambda \) random time
  - Death: with probability \( \frac{#(x)}{#(x) + \int W \lambda^* (x; \xi) d\xi} \), choose a point \( \eta \) of \( x \) at random and delete it from \( x \);
  - Birth: otherwise generate a new point \( \xi \) with probability density proportional to \( \lambda^* (x; \xi) \), and add it to \( x \).

An easy coupling argument shows this chain is geometrically ergodic (in sense of Defn. 10.3) for sensible \( \lambda^* (x; \xi) \).

Note the significant role of Papangelou conditional intensity \( \lambda^* (x; \xi) \) ("chance of a point at \( x \) given rest of pattern \( \xi \)).

3 "geometrically ergodic": achieves equilibrium geometrically fast.
Algorithm to generate a random point pattern $X$ (see, eg, Stoyan et al. 1995). Repeat the following:

- suppose the current value of $X$ is $X = x$:
- after Exponential $(\#(x) + \int_W \lambda^*(x; \xi) \, d\xi)$ random time

$\text{Death}$: with probability $\frac{\#(x)}{\#(x) + \int_W \lambda^*(x; \xi) \, d\xi}$, choose a point $\eta$ of $x$ at random and delete it from $x$;

$\text{Birth}$: otherwise generate a new point $\xi$ with probability density proportional to $\lambda^*(x; \xi)$, and add it to $x$.

An easy coupling argument shows this chain is geometrically ergodic (in sense of Defn. 10.3) for sensible $\lambda^*(x; \xi)$!

Note the significant role of Papangelou conditional intensity $\lambda^*(x; \xi)$ ("chance of a point at $x$ given rest of pattern $\xi$").
Algorithm to generate a random point pattern $X$ (see, eg, Stoyan et al. 1995). Repeat the following:

- suppose the current value of $X$ is $X = x$:
- after Exponential $(\#(x) + \int_W \lambda^*(x; \xi) \, d\xi)$ random time
  - **Death**: with probability $\#(x)/(\#(x) + \int_W \lambda^*(x; \xi) \, d\xi)$, choose a point $\eta$ of $x$ at random and delete it from $x$;
Algorithm to generate a random point pattern $X$ (see, eg, Stoyan et al. 1995). Repeat the following:

- suppose the current value of $X$ is $X = x$:
- after Exponential $(\#(x) + \int_W \lambda^*(x; \xi) \, d\xi)$ random time
  - **Death**: with probability $\#(x)/(\#(x) + \int_W \lambda^*(x; \xi) \, d\xi)$, choose a point $\eta$ of $x$ at random and delete it from $x$;
  - **Birth**: otherwise generate a new point $\xi$ with probability density proportional to $\lambda^*(x; \xi)$, and add it to $x$. 

Note the significant role of Papangelou conditional intensity $\lambda^*(x; \xi)$ ("chance of a point at $x$ given rest of pattern $\xi"$). 

"geometrically ergodic": achieves equilibrium geometrically fast.
Algorithm to generate a random point pattern $X$ (see, eg, Stoyan et al. 1995). Repeat the following:

- suppose the current value of $X$ is $X = x$:
- after Exponential $(\#(x) + \int_W \lambda^*(x; \xi) \, d\xi)$ random time
  - **Death**: with probability $\frac{\#(x)}{\#(x) + \int_W \lambda^*(x; \xi) \, d\xi}$, choose a point $\eta$ of $x$ at random and delete it from $x$;
  - **Birth**: otherwise generate a new point $\xi$ with probability density proportional to $\lambda^*(x; \xi)$, and add it to $x$.

An easy coupling argument shows this chain is geometrically ergodic\(^3\) (in sense of Defn. 10.3) for sensible $\lambda^*(x; \xi)$!

---

\(^3\)“geometrically ergodic”: achieves equilibrium geometrically fast.
Point processes

Algorithm to generate a random point pattern $X$ (see, eg, Stoyan et al. 1995). Repeat the following:

- suppose the current value of $X$ is $X = x$:
- after Exponential $(\#(x) + \int_{W} \lambda^*(x; \xi) \, d\xi)$ random time
  - **Death:** with probability $\#(x)/((\#(x) + \int_{W} \lambda^*(x; \xi) \, d\xi)$, choose a point $\eta$ of $x$ at random and delete it from $x$;
  - **Birth:** otherwise generate a new point $\xi$ with probability density proportional to $\lambda^*(x; \xi)$, and add it to $x$.

An easy coupling argument shows this chain is geometrically ergodic\(^3\) (in sense of Defn. 10.3) for sensible $\lambda^*(x; \xi)$!

Note the significant rôlé of Papangelou conditional intensity $\lambda^*(x; \xi)$ (“chance of a point at $x$ given rest of pattern $\xi$”).

\(^3\)“geometrically ergodic”: achieves equilibrium geometrically fast.
Examples

Strauss process

\[ \lambda^*(x; \xi) = \gamma \text{neighbours of } x \text{ within distance } r \]  \hspace{1cm} (22)

yields density weighting realizations using

\[ \gamma \text{pairs in pattern at distance } r \text{ or less} \].
Examples

Strauss process

$$\lambda^*(x; \xi) = \gamma^\text{neighbours of x within distance } r$$ (22)

yields density weighting realizations using

$$\gamma^\text{pairs in pattern at distance } r \text{ or less}.$$ .

Area-interaction process

$$\lambda^*(x; \xi) = \gamma^\text{area of region within } r \text{ of } x \text{ but not rest of pattern}$$

density \propto \gamma^\text{area of region within distance } r \text{ of pattern}$$ (23)
Domination by a random process

Dominated CFTP replaces the deterministic maximum by a known random process run backwards in time, providing starts for upper- and lower-envelope processes guaranteed to sandwich a valid simulation.
Nonlinear birth and death processes

Suppose $X$ is a nonlinear immigration-death process:

- $X \to X - 1$ at rate $\mu_X$;
- $X \to X + 1 \ldots \alpha_X$ where $\alpha_X \leq \alpha_\infty$.

No maximum (not uniformly ergodic); no classic $CFTP$!
Nonlinear birth and death processes

Suppose $X$ is a nonlinear immigration-death process:

$X \rightarrow X - 1$ at rate $\mu_X$;

$X \rightarrow X + 1 \ldots \alpha_X$ where $\alpha_X \leq \alpha_\infty$.

No maximum (not uniformly ergodic); no classic CFTP!

We can bound by a linear immigration-death process $Y$:

$Y \rightarrow Y - 1$ at rate $\mu_Y$;

$Y \rightarrow Y + 1 \ldots \alpha_\infty$.  


Nonlinear birth and death processes

Suppose $X$ is a nonlinear immigration-death process:

$X \to X - 1$ at rate $\mu X$;
$X \to X + 1 \ldots \alpha_X$ where $\alpha_X \leq \alpha_\infty$.

No maximum (not uniformly ergodic); no classic CFTP! We can bound by a linear immigration-death process $Y$:

$Y \to Y - 1$ at rate $\mu Y$;
$Y \to Y + 1 \ldots \alpha_\infty$.

Produce $X$ from $Y$ by censoring births and deaths:

if $Y \to Y - 1$ then $X \to X - 1$ with c.prob. $X/Y$
if $Y \to Y + 1 \ldots$ $X \to X + 1 \ldots \alpha_X/\alpha_\infty$
Given a trajectory of $Y$, build trajectories of $X$ starting at every $0 \leq X_0 \leq Y_0$ and then staying below $Y$. 
Given a trajectory of $Y$, build trajectories of $X$ starting at every $0 \leq X_0 \leq Y_0$ and then staying below $Y$.

Because $Y$ is *reversible*, with *known equilibrium* (detailed balance!) we can simulate $Y$ backwards, then run forwards with sandwiched $X$ realizations. This is enough for *CFTP*. . .
Domination here can be provided by a constant birth-rate spatial birth and death process. The protocol for the nonlinear immigration-death process can be adapted easily to obtain $\text{domCFTP}$.

Here is an example in which the decision, whether or not to let a birth proceed, is implemented by using Poisson point clusters . . . .

See also Huber (1999)’s notion of a “swap move”. If birth proposal is blocked by just one point, then replace old point by new in a $\text{swap}$, with swap probability $p_{\text{swap}}$ which we are free to choose. Hence “bounding chain”, “sure/not sure” dichotomy.
Fast attractive area-interaction *CFTP*

Häggström, van Lieshout, and Møller (1999) describe fast *CFTP* for attractive area-interaction point processes using special features. If $\gamma > 1$, the density is proportional to the probability that a certain Poisson process places no points within distance $r$ of the pattern. So it may be represented as the pattern of red points, where red and blue points are distributed as Poisson patterns conditioned to be at least distance $r$ from blue and red points respectively. This is monotonic (attractive case only!) and allows Gibbs’ sampler, hence (classic!) *CFTP*.

Gibbs’ sampler *CFTP* for the attractive area-interaction point process as marginal of 2-type soft-core repulsion point process.
Layered Multishift \textit{CFTP}

\textbf{Question:} 
How to draw simultaneously from Uniform($x, x + 1$) for all $x \in \mathbb{R}$, and couple the draws?
Layered Multishift CFTP

**Question:**
How to draw simultaneously from \( \text{Uniform}(x, x + 1) \) for all \( x \in \mathbb{R} \), and couple the draws?

**Answer:** random *unit span integer lattice* Wilson (2000b).
Layered Multishift \textit{CFTP}

Question:
How to draw simultaneously from \( \text{Uniform}(x, x + 1) \) for all \( x \in \mathbb{R} \), and couple the draws?

\textbf{Answer:} random \textit{unit span integer lattice} Wilson (2000b). Consider more general distributions (Wilson 2000b)! For example, we can express a normal distribution as a mixture of uniform distributions, and this allows us to do the same trick in several different ways.
Question:
How to draw simultaneously from Uniform\((x, x + 1)\) for all \(x \in \mathbb{R}\), and couple the draws?

Answer: random unit span integer lattice Wilson (2000b). Consider more general distributions (Wilson 2000b)! For example, we can express a normal distribution as a mixture of uniform distributions, and this allows us to do the same trick in several different ways.
Worked example: \textit{CFTP} for perpetuities

Draw from $\mathcal{L} (X)$ where

$$X \overset{D}{=} U^\alpha (1 + X),$$

for a Uniform($0, 1$) random variable $U$ (and fixed $\alpha > 0$).
Worked example: \textit{CFTP} for perpetuities

Draw from $\mathcal{L}(X)$ where

$$X \overset{D}{=} U^\alpha (1 + X),$$

for a Uniform(0, 1) random variable $U$ (and fixed $\alpha > 0$). This is the \textit{Vervaat} family of perpetuities. Recurrence formulation

$$X_{n+1} = U_{n+1}^\alpha (1 + X_n). \number{24}$$
Implementation details

▶ Dominate on the log-scale (because $\ln(X)$ behaves like a random walk).
Implementation details

- Dominate on the log-scale (because \( \ln(X) \) behaves like a random walk).
- Dominating process \( Y \) behaves like the workload process of an \( M/D/1 \) queue.
- Make life easier: replace \( Y \) by a reflecting simple random walk!
- Impute the continuously-distributed innovations \( U_{n+1} \) of the target process from the discrete jumps of the simple random walk.
- Use multi-shift trick instead of synchronous coupling!
Implementation details

- Dominate on the log-scale (because $\ln(X)$ behaves like a random walk).
- Dominating process $Y$ behaves like the workload process of an $M/D/1$ queue.
- Make life easier: replace $Y$ by a reflecting simple random walk!
Implementation details

- Dominate on the log-scale (because \( \ln(X) \) behaves like a random walk).
- Dominating process \( Y \) behaves like the workload process of an \( M/D/1 \) queue.
- Make life easier: replace \( Y \) by a reflecting simple random walk!
- **Impute** the continuously-distributed innovations \( U_{n+1} \) of the target process from the discrete jumps of the simple random walk.
Implementation details

- Dominate on the log-scale (because $\ln(X)$ behaves like a random walk).
- Dominating process $Y$ behaves like the workload process of an $M/D/1$ queue.
- Make life easier: replace $Y$ by a reflecting simple random walk!
- **Impute** the continuously-distributed innovations $U_{n+1}$ of the target process from the discrete jumps of the simple random walk.
- Use multi-shift trick instead of synchronous coupling!
Perfect simulations of perpetuities

A perfect perpetuity with $\alpha = 1.0$. The *approximate* algorithm, “run recurrence till initial conditions lost in floating point error” is *slower!* 4

---

4The discrepancy will worsen with the move to 64-bit computing.
Perfect simulations of perpetuities

A perfect perpetuity with $\alpha = 1.0$. The *approximate* algorithm, “run recurrence till initial conditions lost in floating point error” is *slower!*  

Another perfect perpetuity with $\alpha = 0.1$. Slower but still very feasible.

*The discrepancy will worsen with the move to 64-bit computing.*
We are all agreed that your theory is crazy. The question which divides us is whether it is crazy enough to have a chance of being correct. My own feeling is that it is not crazy enough.

— Niels Bohr
An important alternative to CFTP makes much fuller use of the notion of time reversal, which we have encountered already in our discussion of queues, and indeed right at the start when we discussed card shuffling.

*The reverse side also has a reverse side.*
— Japanese Proverb

Perfection (FMMR)
  - Siegmund duality
  - FMMR
  - Read-once randomness
  - CFTP for many samples
We begin with a beautiful duality.

**Theorem 9.1**

*Suppose $X$ is a process on $[0, \infty)$. When is there another process $Y$ satisfying the following?*

$$
P[X_t \geq y \mid X_0 = x] = P[Y_t \leq x \mid Y_0 = y] \quad (25)
$$
Siegmund duality

We begin with a beautiful duality.

**Theorem 9.1**

*Suppose* $X$ *is a process on* $[0, \infty)$. *When is there another process* $Y$ *satisfying the following?*

\[ P[X_t \geq y | X_0 = x] = P[Y_t \leq x | Y_0 = y] \quad (25) \]

*Answer (Siegmund 1976): exactly when* $X$ *is (suitably regular and) stochastically monotone:*

\[ P[X_t \geq y | X_0 = x] \leq P[X_t \geq y | X_0 = x'] \quad \text{for} \; x \leq x'. \]
Siegmund duality

We begin with a beautiful duality.

**Theorem 9.1**

*Suppose $X$ is a process on $[0, \infty)$. When is there another process $Y$ satisfying the following?*

$$
P [X_t \geq y | X_0 = x] = P [Y_t \leq x | Y_0 = y] 
$$

(25)

*Answer (Siegmund 1976): exactly when $X$ is (suitably regular and) stochastically monotone:*

$$
P [X_t \geq y | X_0 = x] \leq P [X_t \geq y | X_0 = x'] \text{ for } x \leq x'.
$$

**Proof.**

Derive Chapman-Kolmogorov equations from (25), Fubini.
If $X$ is not stochastically monotone then we get negative transition probabilities for $Y$!

**Consequence:** $Y$ is absorbed at 0, and $X$ at $\infty$. 
If $X$ is not stochastically monotone then we get negative transition probabilities for $Y$!

**Consequence:** $Y$ is absorbed at 0, and $X$ at $\infty$.

Consider simple symmetric random walk on non-negative integers, reflected at 0. Show its Siegmund dual is simple symmetric random walk on non-negative integers, absorbed at 0.
If $X$ is not stochastically monotone then we get negative transition probabilities for $Y$!

**Consequence:** $Y$ is absorbed at 0, and $X$ at $\infty$.

Consider simple symmetric random walk on non-negative integers, reflected at 0. Show its Siegmund dual is simple symmetric random walk on non-negative integers, absorbed at 0.

**Intuition:** think of the Siegmund dual this way. Couple monotonically the $X(x)$ begun at different $x$ (use stochastic monotonicity!), set $Y_t(y) = x$ if $X_0(x) = y$. 

EXERCISE 9.1

EXERCISE 9.2
If $X$ is not stochastically monotone then we get negative transition probabilities for $Y$!

**Consequence:** $Y$ is absorbed at 0, and $X$ at $\infty$.

Consider simple symmetric random walk on non-negative integers, reflected at 0. Show its Siegmund dual is simple symmetric random walk on non-negative integers, absorbed at 0.

**Intuition:** think of the Siegmund dual this way. Couple monotonically the $X^{(x)}$ begun at different $x$ (use stochastic monotonicity!), set $Y^{(y)}_t = x$ if $X^{(x)}_0 = y$.

This beautiful idea grew into a method of simulation, and then a method of perfect simulation (Fill 1998) alternative to CFTP. Fill’s method is harder to explain than CFTP.
If $X$ is not stochastically monotone then we get negative transition probabilities for $Y$!

**Consequence:** $Y$ is absorbed at 0, and $X$ at $\infty$.

Consider simple symmetric random walk on non-negative integers, reflected at 0. Show its Siegmund dual is simple symmetric random walk on non-negative integers, absorbed at 0.

**Intuition:** think of the Siegmund dual this way. Couple monotonically the $X^{(x)}$ begun at different $x$ (use stochastic monotonicity!), set $Y_t^{(y)} = x$ if $X_0^{(x)} = y$.

This beautiful idea grew into a method of simulation, and then a method of perfect simulation (Fill 1998) alternative to CFTP. Fill’s method is harder to explain than CFTP. However we know now how to relate the two, in a very striking way . . . which also makes clear when and how Fill’s method can be advantageous.
The alternative to CFTP is Fill’s algorithm (Fill 1998; Thönnes 1999), at first sight quite different, based on the notion of a strong uniform time \( T \) (Diaconis and Fill 1990) and related to Siegmund duality. Fill et al. (2000) establish a profound link. We explain using “blocks” as input-output maps for a chain.

First recall that CFTP can be viewed in a curiously redundant fashion as follows:

1. Draw from equilibrium \( X(-T) \) and run forwards;
2. continue to increase \( T \) until \( X(0) \) is coalesced;
3. return \( X(0) \).
Key observation: By construction, $X(-T)$ is independent of $X(0)$ and $T$ so . . .

- Condition on a convenient $X(0)$;
- Run $X$ backwards to a fixed time $-T$;
- Draw blocks conditioned on the $X$ transitions;
- If coalescence then return $X(-T)$ else repeat.

Is there a dominated version of Fill's method?

“It's a kind of magic . . .”

Queen
Wilson (2000a) shows how to avoid a conventional requirement of \textit{CFTP}, to re-use randomness used in each cycle. A brief description uses the blocks picture.

- input-output maps from Markov chain coupling;
- Sub-sample to produce “blocks”, positive probability $p$ (say, $p > 1/2$) of block being coalescent;
- \textbf{Algorithm}: repeatedly apply blocks. After first coalescent, return values prior to coalescents.

\textbf{Key}: view simple Small-Set \textit{CFTP} as composition of Geometrically many kernels (conditioned to miss small set).
CFTP for many samples

Mostly people want *many* samples not just one. Murdoch and Rosenthal (2000) consider various strategies for using sample paths produced by *CFTP etc* as effectively as possible:

1. **Repeated CFTP**: repeat CFTP followed by $T_0$ more values.
   - Independence!
2. **Forward coupling**: run forwards to coalescence, then repeat using coalesced value as initial start.
   - Not CFTP!
3. **Concatenated CFTP**: use coalesced value from one CFTP to define path for next CFTP.
   - MA($1$) structure!
4. **Guaranteed time CFTP**: like concatenation, but extract only the last $T_g$ steps of each path.
   - MA($1$) structure!

All the CFTP methods have similar costs if care is taken over parameters: decide whether you need independence! Forward coupling (or other non-CFTP methods) may be better if perfection is not a requirement.

These ideas led rapidly to the FMMR formulation . . .
CFTP for many samples

Mostly people want *many* samples not just one. Murdoch and Rosenthal (2000) consider various strategies for using sample paths produced by *CFTP etc* as effectively as possible:

1. Repeated *CFTP*: repeat *CFTP* followed by $T_0$ more values.
   Independence!

These ideas led rapidly to the FMMR formulation . . . .
CFTP for many samples

Mostly people want *many* samples not just one. Murdoch and Rosenthal (2000) consider various strategies for using sample paths produced by *CFTP etc* as effectively as possible:

1. **Repeated CFTP**: repeat *CFTP* followed by $T_0$ more values. *Independence!*

2. **Forward coupling**: run forwards to coalescence, then repeat using coalesced value as initial start. *Not CFTP!*
CFTP for many samples

Mostly people want many samples not just one. Murdoch and Rosenthal (2000) consider various strategies for using sample paths produced by CFTP etc as effectively as possible:

1. Repeated CFTP: repeat CFTP followed by $T_0$ more values. Independence!

2. Forward coupling: run forwards to coalescence, then repeat using coalesced value as initial start. Not CFTP!

3. Concatenated CFTP: use coalesced value from one CFTP to define path for next CFTP. MA(1) structure!
CFTP for many samples

Mostly people want many samples not just one. Murdoch and Rosenthal (2000) consider various strategies for using sample paths produced by CFTP etc as effectively as possible:

1. Repeated CFTP: repeat CFTP followed by $T_0$ more values. Independence!

2. Forward coupling: run forwards to coalescence, then repeat using coalesced value as initial start. Not CFTP!

3. Concatenated CFTP: use coalesced value from one CFTP to define path for next CFTP. MA(1) structure!

4. Guaranteed time CFTP: like concatenation, but extract only the last $T_g$ steps of each path. MA(1) structure!
CFTP for many samples

Mostly people want many samples not just one. Murdoch and Rosenthal (2000) consider various strategies for using sample paths produced by CFTP etc as effectively as possible:

1. Repeated CFTP: repeat CFTP followed by $T_0$ more values. Independence!

2. Forward coupling: run forwards to coalescence, then repeat using coalesced value as initial start. Not CFTP!

3. Concatenated CFTP: use coalesced value from one CFTP to define path for next CFTP. MA(1) structure!

4. Guaranteed time CFTP: like concatenation, but extract only the last $T_g$ steps of each path. MA(1) structure!

All the CFTP methods have similar costs if care is taken over parameters: decide whether you need independence! Forward coupling (or other non-CFTP methods) may be better if perfection is not a requirement.
CFTP for many samples

Mostly people want *many* samples not just one. Murdoch and Rosenthal (2000) consider various strategies for using sample paths produced by *CFTP etc* as effectively as possible:

1. **Repeated CFTP**: repeat CFTP followed by $T_0$ more values. *Independence!*

2. **Forward coupling**: run forwards to coalescence, then repeat using coalesced value as initial start. *Not CFTP!*

3. **Concatenated CFTP**: use coalesced value from one CFTP to define path for next CFTP. *MA(1) structure!*

4. **Guaranteed time CFTP**: like concatenation, but extract only the last $T_g$ steps of each path. *MA(1) structure!*

All the CFTP methods have similar costs if care is taken over parameters: decide whether you need independence! Forward coupling (or other non-CFTP methods) may be better if perfection is not a requirement. These ideas led rapidly to the FMMR formulation . . . .
There is something tragic about the enormous number of men there are in England at the present moment who start life with perfect profiles, and end by adopting some useful profession.

— Oscar Wilde, Phrases and Philosophies for the use of the Young
Lecture 10: Sundry further topics in \textit{CFTP}

\textit{Stult's Report:}
\begin{quote}
Our problems are mostly behind us.
What we have to do now is fight the solutions.
\end{quote}

Sundry further topics in CFTP
- Price of perfection
- Efficiency
- Impractical CFTP
- Geometric Ergodicity
- Impractical domCFTP theorem
The price of perfection

What kind of price might we have to pay for the perfection of CFTP?
The price of perfection

What kind of price might we have to pay for the perfection of CFTP?

Propp and Wilson (1996) bound for monotonic CFTP on a finite partially ordered space: if \( \ell \) is longest chain, \( T^* \) is coalescence time, and

\[
\overline{d}(k) = \max_{x,y} \{ \text{dist}_{TV}(P^{(k)}_x, P^{(k)}_y) \}.
\]  

(26)

EXERCISE 10.1
The price of perfection

What kind of price might we have to pay for the perfection of CFTP?

Propp and Wilson (1996) bound for monotonic CFTP on a finite partially ordered space: if $\ell$ is longest chain, $T^*$ is coalescence time, and

$$
\overline{d}(k) = \max_{x,y}\{\text{dist}_{TV}(P_x^{(k)}, P_y^{(k)})\}.
$$

(26)

then CFTP is within a factor of being as good as possible:

$$
\frac{\Pr [T^* > k]}{\ell} \leq \overline{d}(k) \leq \Pr [T^* > k].
$$

(27)
(Co-adapted) CFTP is limited by rate of co-adapted coupling: we expect both this and rate of convergence to equilibrium to be exponential.
Efficiency

(Co-adapted) $CFTP$ is limited by rate of co-adapted coupling: we expect both this and rate of convergence to equilibrium to be exponential.

Let $\tau$ be the coupling time, $X = (X_1, X_2)$ the coupled pair.
(Co-adapted) CFTP is limited by rate of co-adapted coupling: we expect both this and rate of convergence to equilibrium to be exponential.

Let $\tau$ be the coupling time, $X = (X_1, X_2)$ the coupled pair. Suppose $|p_t(x_1, y) - p_t(x_2, y)| \approx c_2 \exp(-\mu_2 t)$ while $\mathbb{P}[\tau > t | X(0) = (x_1, x_2)] \approx c \exp(-\mu t)$
Efficiency

(Co-adapted) CFTP is limited by rate of co-adapted coupling: we expect both this and rate of convergence to equilibrium to be exponential.

Let $\tau$ be the coupling time, $X = (X_1, X_2)$ the coupled pair. Suppose $|p_t(x_1, y) - p_t(x_2, y)| \approx c_2 \exp(-\mu_2 t)$ while $\mathbb{P}[\tau > t|X(0) = (x_1, x_2)] \approx c \exp(-\mu t)$

By the coupling inequality, coupling cannot happen faster than convergence. Can it be slower? Burdzy and Kendall (2000) show that the answer is yes.
Coupling of couplings

Letting $X = (X_1, X_2)$, $x = (x_1, x_2)$,

$$|p_t(x_1, y) - p_t(x_2, y)| =$$

$$= |\mathbb{P}[X_1(t) = y | X_1(0) = x_1] - \mathbb{P}[X_2(t) = y | X_2(0) = x_2]|$$

Thus $\mu_2 \geq \mu'_2 + \mu_2$. 
Coupling of couplings

Letting $X = (X_1, X_2)$, $x = (x_1, x_2)$,

$$|p_t(x_1, y) - p_t(x_2, y)| =$$

$$= |P[X_1(t) = y | X_1(0) = x_1] - P[X_2(t) = y | X_2(0) = x_2]| \leq$$

$$|P[X_1(t) = y | \tau > t, X(0) = x] - P[X_2(t) = y | \tau > t, X(0) = x]|$$

$$\times P[\tau > t | X(0) = x]$$

Thus $\mu^2 \geq \mu'_2 + \mu_2$. 
Coupling of couplings

Letting $X = (X_1, X_2), \ x = (x_1, x_2),$

\[
|p_t(x_1, y) - p_t(x_2, y)| =
= |\mathbb{P}[X_1(t) = y | X_1(0) = x_1] - \mathbb{P}[X_2(t) = y | X_2(0) = x_2]| \leq
|\mathbb{P}[X_1(t) = y | \tau > t, X(0) = x] - \mathbb{P}[X_2(t) = y | \tau > t, X(0) = x]| \times \mathbb{P}[\tau > t | X(0) = x]
\]

Let $X^*$ be a coupled copy of $X$ but begun at $(x_2, x_1)$:
Coupling of couplings

Letting $X = (X_1, X_2)$, $x = (x_1, x_2)$,

$$|p_t(x_1, y) - p_t(x_2, y)| =$$

$$= |\mathbb{P}[X_1(t) = y|X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|X_2(0) = x_2]| \leq$$

$$\leq |\mathbb{P}[X_1(t) = y|\tau > t, X(0) = x] - \mathbb{P}[X_2(t) = y|\tau > t, X(0) = x]|$$

$$\times \mathbb{P}[\tau > t|X(0) = x]$$

Let $X^*$ be a coupled copy of $X$ but begun at $(x_2, x_1)$:

$$|\mathbb{P}[X_1(t) = y|\tau > t, X(0) = x] - \mathbb{P}[X_2(t) = y|\tau > t, X(0) = x]|$$

$$= |\mathbb{P}[X_1(t) = y|\tau > t, X(0) = x] -$$

$$- \mathbb{P}[X_1^*(t) = y|\tau > t, X^*(0) = (x_2, x_1)]|$$
Coupling of couplings

Letting $X = (X_1, X_2)$, $x = (x_1, x_2)$,

$$|p_t(x_1, y) - p_t(x_2, y)| =$$

$$= |\mathbb{P}[X_1(t) = y|X_1(0) = x_1] - \mathbb{P}[X_2(t) = y|X_2(0) = x_2]| \leq$$

$$|\mathbb{P}[X_1(t) = y|\tau > t, X(0) = x] - \mathbb{P}[X_2(t) = y|\tau > t, X(0) = x]|$$

$$\times \mathbb{P}[\tau > t|X(0) = x]$$

Let $X^*$ be a coupled copy of $X$ but begun at $(x_2, x_1)$:

$$|\mathbb{P}[X_1(t) = y|\tau > t, X(0) = x] - \mathbb{P}[X_2(t) = y|\tau > t, X(0) = x]|$$

$$= |\mathbb{P}[X_1(t) = y|\tau > t, X(0) = x] -$$

$$\mathbb{P}[X_1^*(t) = y|\tau > t, X^*(0) = (x_2, x_1)]|$$

$$\leq \mathbb{P}[\sigma > t|\tau > t, X(0) = x] \quad (\approx c' \exp(-\mu' t))$$

for $\sigma$ the time when $X$, $X^*$ couple.
Coupling of couplings

Letting $X = (X_1, X_2)$, $x = (x_1, x_2)$,

$$|p_t(x_1, y) - p_t(x_2, y)| =$$

$$= \left| \mathbb{P}[X_1(t) = y | X_1(0) = x_1] - \mathbb{P}[X_2(t) = y | X_2(0) = x_2] \right| \leq$$

$$|\mathbb{P}[X_1(t) = y | \tau > t, X(0) = x] - \mathbb{P}[X_2(t) = y | \tau > t, X(0) = x]|$$

$$\times \mathbb{P}[\tau > t | X(0) = x]$$

Let $X^*$ be a coupled copy of $X$ but begun at $(x_2, x_1)$:

$$|\mathbb{P}[X_1(t) = y | \tau > t, X(0) = x] - \mathbb{P}[X_2(t) = y | \tau > t, X(0) = x]| =$$

$$= |\mathbb{P}[X_1(t) = y | \tau > t, X(0) = x] -$$

$$\mathbb{P}[X^*_1(t) = y | \tau > t, X^*(0) = (x_2, x_1)]|$$

$$\leq \mathbb{P}[\sigma > t | \tau > t, X(0) = x] \quad (\approx c' \exp(-\mu' t))$$

for $\sigma$ the time when $X, X^*$ couple. Thus $\mu_2 \geq \mu' + \mu$. 
Impractical *CFTP*

When can we *conceive* of being able to do classic *CFTP*?
Impractical CFTP

When can we conceive of being able to do classic CFTP?

Theorem 10.1 (Foss and Tweedie 1998)

*(Impractical) classic CFTP is equivalent to uniform ergodicity.*

\[ \text{(Impractical) classic CFTP is equivalent to uniform ergodicity.} \]
Impractical $CFTP$

When can we conceive of being able to do classic $CFTP$?

**Theorem 10.1** (Foss and Tweedie 1998)

(Impractical) classic $CFTP$ is equivalent to uniform ergodicity.

**Definition 10.2** (Uniform Ergodicity)

The chain $X$ is uniformly ergodic, equilibrium distribution $\pi$, if there is a constant $V$ and $\gamma \in (0, 1)$,

$$\| P^n(x, \cdot) - \pi(\cdot) \|_{TV} \leq V \gamma^n \quad \text{for all } n, x. \quad (28)$$
Impractical \textit{CFTP}

When can we conceive of being able to do classic \textit{CFTP}?

\textbf{Theorem 10.1 (Foss and Tweedie 1998)} \\
(Impractical) classic \textit{CFTP} is equivalent to uniform ergodicity.

\textbf{Definition 10.2 (Uniform Ergodicity)} \\
The chain $X$ is uniformly ergodic, equilibrium distribution $\pi$, if there is a constant $V$ and $\gamma \in (0, 1)$,

$$\| P^{(n)}(x, \cdot) - \pi(\cdot) \|_{TV} \leq V \gamma^n \quad \text{for all } n, x.$$  \hspace{1cm} (28)

Hint: uniform ergodicity is \textbf{equivalent} to the entire state-space being small, therefore one can use small-set \textit{CFTP}!
Impractical CFTP

When can we conceive of being able to do classic CFTP?

Theorem 10.1 (Foss and Tweedie 1998) (Impractical) classic CFTP is equivalent to uniform ergodicity.

Definition 10.2 (Uniform Ergodicity)
The chain $X$ is uniformly ergodic, equilibrium distribution $\pi$, if there is a constant $V$ and $\gamma \in (0, 1)$,

$$\| P^n(x, \cdot) - \pi(\cdot) \|_{TV} \leq V \gamma^n$$

for all $n, x$. (28)

Hint: uniform ergodicity is equivalent to the entire state-space being small, therefore one can use small-set CFTP!

So what is the scope of $\text{domCFTP}$? How far can we go?
Definition 10.3 (Geometric Ergodicity)

The chain $X$ is geometrically ergodic, equilibrium distribution $\pi$, if there is a $\pi$-almost surely finite $V : \mathcal{X} \to [0, \infty]$ and $\gamma \in (0, 1)$,

$$\| P^{(n)}(x, \cdot) - \pi(\cdot) \|_{TV} \leq V(x) \gamma^n \quad \text{for all } n, x.$$  (29)
Geometric Ergodicity

Definition 10.3 (Geometric Ergodicity)
The chain $X$ is geometrically ergodic, equilibrium distribution $\pi$, if there is a $\pi$-almost surely finite $V : \mathcal{X} \to [0, \infty]$ and $\gamma \in (0, 1)$,

$$\|P^{(n)}(x, \cdot) - \pi(\cdot)\|_{TV} \leq V(x) \gamma^n \quad \text{for all } n, x . \quad (29)$$

Equivalently
Geometric Ergodicity

Definition 10.3 (Geometric Ergodicity)
The chain $X$ is geometrically ergodic, equilibrium distribution $\pi$, if there is a $\pi$-almost surely finite $V : \mathcal{X} \to [0, \infty]$ and $\gamma \in (0, 1)$,

$$\|P^{(n)}(x, \cdot) - \pi(\cdot)\|_{TV} \leq V(x)\gamma^n \quad \text{for all } n, x.$$  \hspace{1cm} (29)

Equivalently

Definition 10.4 (Geometric Foster-Lyapunov Criterion)
There are: a small set $C$; constants $\alpha \in (0, 1), b > 0$; and a scale or drift function $\Lambda : \mathcal{X} \to [1, \infty)$ bounded on $C$; such that for $\pi$-almost all $x \in \mathcal{X}$

$$\mathbb{E}[\Lambda(X_{n+1}) | X_n = x] \leq \alpha\Lambda(x) + b\mathbb{I}[x \in C].$$ \hspace{1cm} (30)
Geometric Ergodicity

Definition 10.3 (Geometric Ergodicity)
The chain $X$ is geometrically ergodic, equilibrium distribution $\pi$, if there is a $\pi$-almost surely finite $V : \mathcal{X} \to [0, \infty]$ and $\gamma \in (0, 1)$,

$$\|P^{(n)}(x, \cdot) - \pi(\cdot)\|_{TV} \leq V(x) \gamma^n$$

for all $n, x$. (29)

Equivalently

Definition 10.4 (Geometric Foster-Lyapunov Criterion)
There are: a small set $C$; constants $\alpha \in (0, 1), b > 0$; and a scale or drift function $\Lambda : \mathcal{X} \to [1, \infty)$ bounded on $C$; such that for $\pi$-almost all $x \in \mathcal{X}$

$$\mathbb{E} [\Lambda(X_{n+1}) \mid X_n = x] \leq \alpha \Lambda(x) + b \mathbb{I}[x \in C].$$

(30)

... and there are still other kinds of ergodicity!
Impractical domCFTP theorem

Theorem 10.5 (Kendall 2004)

(Impractical) domCFTP is possible in principle for any geometrically ergodic Markov chain.
Impractical *domCFTP* theorem

**Theorem 10.5 (Kendall 2004)**

(Impractical) *domCFTP* is possible in principle for any geometrically ergodic Markov chain.

**IDEA:** try to use Equation (30)

\[
E \left[ \Lambda(X_{n+1}) \mid X_n = x \right] \leq \alpha \Lambda(x) + b \mathbb{I} [x \in C].
\]

to determine a dominating process \( Y \), dominating in the sense that

\[
X \preceq Y \quad \text{if and only if} \quad \Lambda(X) \leq Y.
\]

The two sets of ingredients of the proof run as follows . . .
Use Markov’s inequality and coupling to build $Y$ so if $\Lambda(x_0) \leq y_0$

$$P[\Lambda(X_{n+1}) \geq y \mid X_n = x_0] \leq P[Y_{n+1} \geq y \mid Y_n = y_0].$$
Use Markov’s inequality and coupling to build $Y$ so if $\Lambda(x_0) \leq y_0$

$$\mathbb{P} [ \Lambda(X_{n+1}) \geq y \mid X_n = x_0] \leq \mathbb{P} [ Y_{n+1} \geq y \mid Y_n = y_0] .$$

Deduce from geometric Foster-Lyapunov criterion that

$$\mathbb{P} [ Y_{n+1} \geq y \mid Y_n = y_0] = \frac{\alpha y_0}{y}$$

whenever $y > \alpha y_0$, $y > \max\{\Lambda|C\} + b/\alpha$. 
Ingredients I

Use Markov’s inequality and coupling to build $Y$ so if $\Lambda(x_0) \leq y_0$

$$P[\Lambda(X_{n+1}) \geq y \mid X_n = x_0] \leq P[Y_{n+1} \geq y \mid Y_n = y_0].$$

Deduce from geometric Foster-Lyapunov criterion that

$$P[Y_{n+1} \geq y \mid Y_n = y_0] = \alpha y_0 / y$$

whenever $y > \alpha y_0$, $y > \max\{\Lambda \mid C\} + b / \alpha$,

Add a lower reflecting barrier for $Y$: then $\ln(Y)$ can be related to system-workload for a $D/M/1$ queue, Exponential(1) service times and arrival intervals $\ln(1 / \alpha)$. 
Ingredients II

- If $e^{-1} \leq \alpha \leq 1$ then $D/M/1$ queue is not positive-recurrent!
Ingredients II

- If $e^{-1} \leq \alpha \leq 1$ then $D/M/1$ queue is not positive-recurrent!
- Fix this by sub-sampling (geometric Foster-Lyapunov condition is essentially unaltered by this).

Simple lemma: coupling of $X$, $Y$ persists through small-set regeneration.
Ingredients II

- If $e^{-1} \leq \alpha \leq 1$ then $D/M/1$ queue is not positive-recurrent!
- Fix this by sub-sampling (geometric Foster-Lyapunov condition is essentially unaltered by this).
- Exact formulae for the $D/M/1$ equilibrium and dual process transition kernel allow us to obtain explicit, simple, simulation algorithms!
If $e^{-1} \leq \alpha \leq 1$ then $D/M/1$ queue is not positive-recurrent!

Fix this by sub-sampling (geometric Foster-Lyapunov condition is essentially unaltered by this).

Exact formulae for the $D/M/1$ equilibrium and dual process transition kernel allow us to obtain explicit, simple, simulation algorithms!

Run $D/M/1$ queue backwards in time till it hits barrier, which defines a small set for target chains. (Need to sub-sample again so that this small set is of order 1.)
If $e^{-1} \leq \alpha \leq 1$ then $D/M/1$ queue is not positive-recurrent!

Fix this by sub-sampling (geometric Foster-Lyapunov condition is essentially unaltered by this).

Exact formulae for the $D/M/1$ equilibrium and dual process transition kernel allow us to obtain explicit, simple, simulation algorithms!

Run $D/M/1$ queue backwards in time till it hits barrier, which defines a small set for target chains. (Need to sub-sample again so that this small set is of order 1.)

Simple lemma: coupling of $X, Y$ persists through small-set regeneration.
Impractical domCFTP: conclusion

- This kind of *domCFTP* may be un-implementable for any of several reasons:
Impractical domCFTP: conclusion

- This kind of domCFTP may be un-implementable for any of several reasons:
  - use of sub-sampling: implies substantial knowledge of higher-order transition probabilities!
Impractical domCFTP: conclusion

- This kind of *domCFTP* may be un-implementable for any of several reasons:
  - use of sub-sampling: implies substantial knowledge of higher-order transition probabilities!
  - how to implement the Markov inequality coupling? (especially in case of sub-sampling!)
This kind of domCFTP may be un-implementable for any of several reasons:

- use of sub-sampling: implies substantial knowledge of higher-order transition probabilities!
- how to implement the Markov inequality coupling? (especially in case of sub-sampling!)
- depends on a Foster-Lyapunov condition which may produce a coupling which is much weaker than a more sophisticated approach.
Impractical domCFTP: conclusion

- This kind of *domCFTP* may be un-implementable for any of several reasons:
  - use of sub-sampling: implies substantial knowledge of higher-order transition probabilities!
  - how to implement the Markov inequality coupling? (especially in case of sub-sampling!)
  - depends on a Foster-Lyapunov condition which may produce a coupling which is much weaker than a more sophisticated approach.

- But it is closely linked to applications of *domCFTP* which actually work, such as the perpetuities example in §8.5.
Conclusion

*We can predict everything, except the future.*