Coupling and convergence

Saul Jacka, Warwick Statistics

OxWaSP, Warwick

7 November, 2014

Qn: What is coupling for?
A: To compare probabilities.

Three uses: *ordering*, *approximation* and *convergence* of probabilities.

They all work by creating copies of random objects which simultaneously live on the same probability space and have a desirable relationship. OxWaSP students have already met the idea with the Chen-Stein method.

Qn.: How do we simulate from an arbitrary distribution on $\mathbb{R}$?

Answer: suppose the relevant distribution function is F. Take U, a $U[0,1]$ random variable [under the probability measure $\mathbb{P}$], and set

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

(where $F^{-1}(t) \stackrel{def}{=} \inf\{x : F(x) \geq t\}$ for $t \in [0,1]$).

Note: F is right-continuous and increasing so $F(F^{-1}(t) \geq t$ for all t.

Check:

$$\mathbb{P}(X \leq a) = \mathbb{P}(F^{-1}(U) \leq a) = \mathbb{P}(F(a) \geq U) = F(a), \text{ since } U \text{ is uniform,}$$

so X has distribution function F as required.

Recall that the distribution under $\mathbb{P}$ of a random variable X (defined on $(\Omega, \mathcal{F})$ and taking values in $(E, \mathcal{E})$) is $\mathbb{P}_X$ given by

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

Theorem: (Skorkohod, Dudley: Skorokhod representation)
Suppose that E is a separable metric space (e.g. $\mathbb{R}^n$) with Borel σ-algebra $\mathcal{E}$. Suppose that $(\mathbb{P}_n)_{n \leq \infty}$ are probability measures on $(E, \mathcal{E})$ with $\mathbb{P}_n \overset{w}{\Rightarrow} \mathbb{P}_\infty$, then there exist random objects $(X_n)_{n \leq \infty}$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that

(i) $\mathbb{P}_{X_n} = \mathbb{P}_n \quad n \leq \infty$

and

(ii) $X_n \xrightarrow{a.s.} X_\infty$.

Proof: (in the real-valued case) *Use the corresponding distribution functions* $(F_n)_{n \leq \infty}$*. As before, construct a uniform r.v.* U *and then set*

$$X_n = F_n^{-1}(U).$$

It is (fairly) clear that $F_n^{-1}(t) \to F_\infty^{-1}(t)$ *for all but countably many* t *and so* $X_n \xrightarrow{a.s.} X_\infty$*.*

Theorem: (Radon-Nikodym Theorem) *Suppose that $\mathbb{M}$ and R are σ-finite measures on $(\Omega, \mathcal{F})$ and whenever $R(A) = 0$, $\mathbb{M}(A) = 0$, then we write*

$$\mathbb{M} << R$$

and there exists an $f \geq 0$, such that

$$\int X d\mathbb{M} = \int X f dR, \text{ for all } \mathbb{M}\text{-integrable } X.$$

We often write $f = \frac{d\mathbb{M}}{dR}$ and note that it satisfies the chain rule: $\mathbb{M} << \mathbb{N} << R \Rightarrow \frac{d\mathbb{M}}{dR} = \frac{d\mathbb{M}}{d\mathbb{N}}\frac{d\mathbb{N}}{dR}$. If $\mathbb{M} << R$ and $R << \mathbb{M}$ we write $\mathbb{M} \sim R$.

Exercise:[Ex1] Show that $\frac{d(\mathbb{M}+\mathbb{N})}{dR} = \frac{d\mathbb{M}}{dR} + \frac{d\mathbb{M}}{dR}$ and deduce that if $\mathbb{M} \leq \mathbb{N}$ (i.e. $\mathbb{M}(A) \leq \mathbb{N}(A)$ for all $A \in \mathcal{F}$) then $\frac{d\mathbb{M}}{dR} \leq \frac{d\mathbb{N}}{dR}$.

Definition: *A family of measures* $\{\mu^\theta; \theta \in \Theta\}$ *is said to be* dominated *by a measure* μ *if*

$$\mu^\theta \ll \mu \text{ for all } \theta \in \Theta,$$

and such a μ *is said to be a dominating measure for the family.*

Remark: *Note that for any countable collection* $\{\mathbb{P}_n\}$ *of probability measures on* $(\Omega, \mathcal{F})$ *there is a dominating measure (call it* R*) such that each* $\mathbb{P}_n$ *is absolutely continuous with respect to* R *(and thus has a density by the Radon-Nikodym theorem). To see this simply set*

$$R = \frac{1}{2}(\mathbb{P}_\infty + \sum_{n=1}^{\infty} 2^{-n}\mathbb{P}_n) \tag{1}$$

Note that, in fact, R *is a probability measure.*

Theorem:[Coupling] (The fundamental inequality of coupling) *If* $\mathbb{P}$ *and* $\mathbb{Q}$ *are two probability measures on* $(\Omega, \mathcal{F})$ *then*

(a) *if* X *and* Y *are random objects:* $X, Y : (\Omega', \mathcal{F}', \mathbb{P}') \to (\Omega, \mathcal{F})$ *with distributions* $\mathbb{P}'_X = \mathbb{P}$ *and* $\mathbb{P}'_Y = \mathbb{Q}$ *then*

$$\mathbb{P}'(X \neq Y) \geq d(\mathbb{P}, \mathbb{Q}) \stackrel{def}{=} \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)| \, ;$$

(b) *there exists a probability space* $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ *and random objects* $X, Y : (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) \to (\Omega, \mathcal{F})$ *such that*

(i) $\tilde{\mathbb{P}}_X = \mathbb{P}$ *and* $\tilde{\mathbb{P}}_Y = \mathbb{Q}$

and

(ii) $\tilde{\mathbb{P}}(X \neq Y) = d(\mathbb{P}, \mathbb{Q})$.

Proof: *the proof of this result relies on the following observation: suppose* $\mathbb{P}$ *and* $\mathbb{Q}$ *are as above then, taking any dominating measure* R *(i.e. an* R *s.t.* $\mathbb{P}, \mathbb{Q} \ll R$*), and, defining* $f_{\mathbb{P}} = \frac{d\mathbb{P}}{dR}, f_{\mathbb{Q}} = \frac{d\mathbb{Q}}{dR}$,

$$d(P, \mathbb{Q}) = \int (f_{\mathbb{P}} - f_{\mathbb{Q}})^{+} dR = \int (f_{\mathbb{P}} - f_{\mathbb{Q}})^{-} dR$$

$$\text{(since } \textstyle\int (f_{\mathbb{P}} - f_{\mathbb{Q}}) dR = 0)$$

$$= \int_K (f_{\mathbb{P}} - f_{\mathbb{Q}}) dR = \int_{K^c} (f_{\mathbb{Q}} - f_{\mathbb{P}}) dR = \mathbb{P}(K) - \mathbb{Q}(K) = \mathbb{Q}(K^c) - \mathbb{P}(K^c)$$

where $K = \{\omega : f_{\mathbb{P}}(\omega) > f_{\mathbb{Q}}(\omega)\}$

To prove (a), define the measure S by $S(A) = \mathbb{P}'(X = Y \in A)$ and observe that $S \leq \mathbb{P}$ and $S \leq \mathbb{Q}$ and so (by Exercise [Ex1]) S has density dominated by $f_{\mathbb{P}} \wedge f_{\mathbb{Q}}$. Thus

$$1 - S(\Omega) = \mathbb{P}'(X \neq Y) \geq 1 - \int_\Omega f_{\mathbb{P}} \wedge f_{\mathbb{Q}} dR = \int_\Omega (f_{\mathbb{P}} - f_{\mathbb{P}} \wedge f_{\mathbb{Q}}) dR$$

$$= \int_\Omega (f_{\mathbb{P}} - f_{\mathbb{Q}})^+ dR = d(\mathbb{P}, \mathbb{Q}).$$

To prove (b) we construct independent random variables X, Z and U on the measurable space $(\tilde{\Omega}, \tilde{\mathcal{F}}) \stackrel{def}{=} (\Omega \times \Omega \times [0,1], \mathcal{F} \otimes \mathcal{F} \otimes \mathcal{B}[0,1])$, with U being Uniform$[0,1]$, X having distribution $\mathbb{P}$ and Z having density (wrt R) $f_{\mathbb{M}} \stackrel{def}{=} \frac{(f_{\mathbb{Q}} - f_{\mathbb{P}}) 1_{K^c}}{d(\mathbb{P}, \mathbb{Q})}$.

Thus the probability measure on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ is $\tilde{\mathbb{P}} = \mathbb{P} \otimes \mathbb{M} \otimes \Lambda$, where Λ is Lebesgue measure (on $\mathcal{B}[0,1]$) and, if $\tilde{\omega} = (\omega_1, \omega_2, t)$, then $(X(\tilde{\omega}), Z(\tilde{\omega}), U(\tilde{\omega})) = (\omega_1, \omega_2, t)$, i.e. (X, Z, U) is the identity on $\tilde{\Omega}$.

Now define $Y(\tilde{\omega}) = X 1_{\left(U \leq \frac{f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1)}{f_{\mathbb{P}}(\omega_1)}\right)} + Z 1_{\left(U > \frac{f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1)}{f_{\mathbb{P}}(\omega_1)}\right)}$. Notice that Z takes values in K^c, where $\frac{f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1)}{f_{\mathbb{P}}(\omega_1)} = 1$.

It's clear that X has the right distribution and that

$$\tilde{\mathbb{P}}(Y \in A) = \tilde{\mathbb{P}}(Y = X \in A) + \tilde{\mathbb{P}}(Y = Z \in A)$$

$$= \int_A \frac{f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1)}{f_{\mathbb{P}}(\omega_1)} d\mathbb{P}(\omega_1)$$
$$+ \int_{(\omega_1 \in \Omega)} \int_{(\omega_2 \in A)} \left(1 - \frac{f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1)}{f_{\mathbb{P}}(\omega_1)}\right) d\mathbb{P}(\omega_1) d\mathbb{M}(\omega_2)$$

$$= \int_A f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1) dR + \int_{(\omega_1 \in \Omega)} (f_{\mathbb{P}} - f_{\mathbb{Q}})^+ dR(\omega_1) \int_{(\omega_2 \in A)} d\mathbb{M}(\omega_2)$$

$$= \int_A f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1) dR + d(\mathbb{P}, \mathbb{Q}) \int_{(\omega_2 \in A)} d\mathbb{M}(\omega_2)$$

$$= \int_A f_{\mathbb{P}} \wedge f_{\mathbb{Q}}(\omega_1) dR + \int_{(\omega_2 \in A)} (f_{\mathbb{Q}} - f_{\mathbb{P}})^+ dR(\omega_2) = \int_A f_{\mathbb{Q}} dR = \mathbb{Q}(A),$$

so that Y has distribution $\mathbb{Q}$, as required. Finally, it is clear that $\tilde{\mathbb{P}}(X = Y) = \int f_{\mathbb{P}} \wedge f_{\mathbb{Q}} dR = 1 - d(\mathbb{P}, \mathbb{Q})$, □

Definition: Given a sequence of probability measures $(\mathbb{P}_n)$ we say the $\mathbb{P}_n$ converge **Skorokhod weakly** *to $\mathbb{P}_\infty$, written*

$$\mathbb{P}_n \overset{Sw}{\Rightarrow} \mathbb{P}_\infty \,,$$

if there is a dominating (probability) measure $\mathbb{Q}$ such that:

$$f_n \overset{prob(\mathbb{Q})}{\longrightarrow} f_\infty \text{ as } n \to \infty,$$

where f_n is a version of $\frac{d\mathbb{P}_n}{d\mathbb{Q}}$.

Definition: Given the $(\mathbb{P}_n)_{n\leq\infty}$, we say the $\mathbb{P}_n$ converge **Skorokhod strongly** to $\mathbb{P}_\infty$, written

$$\mathbb{P}_n \overset{Ss}{\Rightarrow} \mathbb{P}_\infty \,,$$

if there exists a dominating probability measure $\mathbb{Q}$ such that:

$$f_\infty \wedge f_n \overset{\mathbb{Q}a.s.}{\longrightarrow} f_\infty \text{ as } n \to \infty \,.$$

Remark: The reason for the nomenclature will become apparent soon.

Remark: There is no need to restrict the choice of $\mathbb{Q}$ to probability measures—any σ-finite measure will do.

Conversely, we gain nothing by allowing more general σ-finite measures, since if R is a σ-finite dominating measure with $T_n \uparrow \Omega$ and $R(T_n) < \infty$ for each n, then there exists a sequence $(a_n) \subset (0, \infty)$ such that $\sum_n a_n R(T_n \backslash T_{n-1}) = 1$ and, defining $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{dR} = \sum_n a_n 1_{(T_n \backslash T_{n-1})},$$

we see that $\mathbb{Q}$ is a probability measure on $(\Omega, \mathcal{F})$ and $\mathbb{Q} \sim R$, so we may substitute $\mathbb{Q}$ for R and $\frac{d\mathbb{P}_n}{d\mathbb{Q}}$ $(\equiv \frac{d\mathbb{P}_n}{dR}\frac{dR}{d\mathbb{Q}})$ for $\frac{d\mathbb{P}_n}{dR}$.

Theorem:[Sw] *if* $(\mathbb{P}_n)$ *are probability measures on* $(\Omega, \mathcal{F})$ *then the following are equivalent*

(i) $\mathbb{P}_n \overset{Sw}{\Rightarrow} \mathbb{P}_\infty$

(ii) $\exists$ *a probability space* $(\Omega', \mathcal{F}', \mathbb{P}')$ *and random objects* $(X_n) : (\Omega', \mathcal{F}', \mathbb{P}') \to (\Omega, \mathcal{F})$ *such that*

   (a) $\mathbb{P}'_{X_n} = \mathbb{P}_n$
   and

   (b) $\mathbb{P}'(X_n \neq X_\infty) \to 0$ *as* $n \to \infty$

(iii) $\mathbb{P}_n \to \mathbb{P}_\infty$ *with respect to the total variation metric i.e.*

$$d(\mathbb{P}_n, \mathbb{P}_\infty) \to 0 \text{ as } n \to \infty$$

(iv) $\exists$ *a dominating probability measure* $\mathbb{Q}$ *s.t. the densities* $f_n = \frac{d\mathbb{P}_n}{d\mathbb{Q}}$ *satisfy*

$$f_n \overset{L^1(\mathbb{Q})}{\longrightarrow} f_\infty$$

(v) $\mathbb{P}_n(A) \to \mathbb{P}_\infty(A)$ *uniformly in* $A \in \mathcal{F}$.

Remark: The equivalence of (i) and (iii) is Scheffé's lemma (see Billingsley (1968)).

Proof: throughout the proof R is a dominating measure.

(iv) $\Rightarrow$ (ii) This mimics part of the proof of the fundamental inequality for coupling. Given $\mathbb{Q}$ and the densities $(f_n)_{n\leq\infty}$, define

$$\begin{aligned}\Omega' &= \Omega \times \Omega^{\infty} \times [0,1],\\ \mathcal{F}' &= \mathcal{F} \otimes \mathcal{F}^{*\infty} \otimes \mathcal{B}([0,1]),\end{aligned}$$

and the probability measures $\mathbb{M}_n$ *by*

$$\frac{d\mathbb{M}_n}{d\mathbb{Q}} = \frac{(f_n - f_\infty)^+}{d(\mathbb{P}_n, \mathbb{P}_\infty)}.$$

Then define

$$\mathbb{P}' = \mathbb{P}_\infty \otimes \bigotimes_{n=1}^{\infty} \mathbb{M}_n \otimes \Lambda$$

and define, for each $\omega' = (\omega_\infty, \omega_1, \ldots; t) \in \Omega'$,

$$\begin{aligned} X_\infty(\omega') = &\quad \omega_\infty, \\ X_n(\omega') = &\quad \omega_\infty \quad 1_{\left(t \le \frac{f_\infty \wedge f_n}{f_\infty}(\omega_1)\right)} + \omega_n 1_{\left(t > \frac{f_\infty \wedge f_n}{f_\infty}(\omega_1)\right)}, \end{aligned}$$

and

$$Y(\omega') = \quad t \quad .$$

What we're doing is making all the coupling constructions simultaneously by constructing X_∞ to have the right law under $\mathbb{P}'$; then, taking a single independent $U[0,1]$ r.v. (called U), setting $X_n = X_\infty$ if (and only if) $U \le \frac{f_\infty \wedge f_n}{f_n}(X_\infty)$ and otherwise giving X_n a conditional distribution which gives it the right (unconditional) distribution. It's not hard to check that $\mathbb{P}'_{X_n} = \mathbb{P}_n$ for all n, whilst

$$\begin{aligned} \mathbb{P}'(X_n \neq X_\infty) &\le (=)\mathbb{P}'\big(U > \tfrac{f_\infty \wedge f_n}{f_\infty}(X_\infty)\big) \\ &= \int_\Omega \frac{(f_n - f_\infty)^+}{f_\infty}\, d\mathbb{P}_\infty \\ &= \int_\Omega (f_n - f_\infty)^+ d\mathbb{Q}, \end{aligned} \tag{2}$$

and by (iv) the last term in (2) tends to 0.

(i) ⇔ (iv) The reverse implication is obvious (since convergence in L^1 *is equivalent to* $\{$*convergence in probability* **and** *uniform integrability*$\}$*). The forward implication follows since (by virtue of the fact that* f_∞ *and* f_n *are densities):*

$$\int_\Omega |f_\infty - f_n| d\mathbb{Q} = 2 \int_\Omega (f_\infty - f_n)^+ d\mathbb{Q} \tag{3}$$

and the integrand on the right of (3) is uniformly bounded by f_∞ *(which is, by definition, in* $L^1(\mathbb{Q})$*).*

(ii) ⇒ (iii) This follows immediately from the coupling inequality.

(iii) $\Rightarrow$ (iv) This follows on taking the dominating measure R:

$$d(\mathbb{P}_n, \mathbb{P}_\infty) \to 0$$

tells us that, letting densities with respect to R be denoted by $f_\cdot^R$,

$$\int_\Omega (f_\infty^R - f_n^R)^+ dR \to 0$$

and (as before) $\int_\Omega |f_\infty^R - f_n^R| dR = 2 \int_\Omega (f_\infty^R - f_n^R)^+ dR$ establishing (iv).

(iii) $\Leftrightarrow$ (v) This is obvious $\square$

Theorem: *[Ss] Suppose* $(\mathbb{P}_n)$ *are probability measures on* $(\Omega, \mathcal{F})$*, then the following are equivalent*

(i) $\mathbb{P}_n \overset{Ss}{\Rightarrow} \mathbb{P}_\infty$

(ii) *There exists a probability space* $(\Omega', \mathcal{F}', \mathbb{P}')$ *and random objects*

$$(X_n) : (\Omega', \mathcal{F}', \mathbb{P}') \to (\Omega, \mathcal{F})$$

such that

(a) $\mathbb{P}'_{X_n} = \mathbb{P}_n$

and

(b) $\mathbb{P}'(X_n \neq X_\infty \text{ i.o.}) = 0.$

Proof: (i) ⇒ (ii) Take the representation given in the proof of the previous theorem, then

$$\mathbb{P}'(\exists n \geq N : X_n \neq X_\infty) = \mathbb{P}'(Y > \inf_{n \geq N} \tfrac{f_\infty \wedge f_n}{f_\infty}(X_\infty))$$
$$= \int_\Omega (1 - \frac{f_\infty \wedge \inf_{n \geq N} f_n}{f_\infty}(\omega))\, d\mathbb{P}_\infty(\omega)$$
$$= \int_\Omega (f_\infty(\omega) - \inf_{n \geq N} f_n(\omega))^+ d\mathbb{Q}(\omega)$$

and by monotone convergence this expression converges to

$$\int_\Omega (f_\infty - \liminf f_n)^+ \, d\mathbb{P}_\infty$$
$$= 0 \text{ (by (i))}.$$

(ii) $\Rightarrow$ (i) Given $\mathbb{P}'$ and (X_n) as in (ii), define $\mathbb{Q}$ as in equation 1, and define, for each $m \geq 1$, the measure $\mathcal{S}_m$ on $(\Omega, \mathcal{F})$ by

$$\mathcal{S}_m(A) = \mathbb{P}'(\exists n \geq m : \ X_n \neq X_\infty, X_\infty \in A),$$

(since $\mathcal{S}_m(A) \leq \mathbb{P}'(X_\infty \in \ A) = \mathbb{P}_\infty(A)$ by hypothesis) $\mathcal{S}_m \ll \mathbb{P}_\infty$, whilst

$$\mathcal{S}_m(\Omega) = \mathbb{P}'(\exists n \geq m : \ X_n \neq X_\infty),$$

so that

$$\lim \mathcal{S}_m(\Omega) = \mathbb{P}'(X_n \neq X_\infty \text{ i.o.}).$$

Now

$$\begin{aligned}
\mathcal{S}_m(A) &\geq \mathbb{P}'(X_n \neq X_\infty, X_\infty \in A) \\
&\geq \mathbb{P}'(X_n \in A^c, X_\infty \in A) \\
&\geq \mathbb{P}'(X_\infty \in A) - \mathbb{P}'(X_n \in A) \\
&= \mathbb{P}_\infty(A) - \mathbb{P}_n(A) \text{ (for any } n \geq m\text{)},
\end{aligned}$$

so that, for any $n \geq m$,

$$g_m \stackrel{def}{=} \frac{d\mathcal{S}_m}{d\mathbb{Q}} \geq f_\infty - f_n \quad (\mathbb{Q} \text{ a.s.}),$$

so

$$g_m \geq (f_\infty - f_n)^+ \quad (\mathbb{Q} \text{ a.s.}) \text{ for any } n \geq m\,.$$

It follows that $g_m \geq (f_\infty - \inf_{n \geq m} f_n)^+$ $(\mathbb{Q}$ *a.s.) and hence*

$$\begin{aligned} 0 &= \lim \mathcal{S}_m(\Omega) = \lim \int_\Omega g_m d\mathbb{Q} \\ &\geq \lim \int_\Omega (f_\infty - \inf_{n \geq m} f_n)^+ d\mathbb{Q}. \end{aligned}$$

It follows (by monotone convergence) that lim inf $f_n \geq f_\infty$ $(\mathbb{Q}$ *a.s.) from which we may easily deduce (using Fatou's lemma) that lim inf* $f_n = f_\infty$ $(\mathbb{Q}$ *a.s.) and hence*

$$f_\infty^{\mathbb{Q}} \wedge f_n \xrightarrow{\mathbb{Q} a.s.} f_\infty$$

□

Exercise:[Ex2] Using counting measure on the integers as a reference measure show that if $(\mathbb{P}_n)_{n\leq\infty}$ are all measures on the integers with $\mathbb{P}_n \overset{w}{\Rightarrow} \mathbb{P}_\infty$, then

$$\mathbb{P}_n \overset{Ss}{\Rightarrow} \mathbb{P}_\infty.$$

Counterexamples In the examples (B_n) are a sequence of Bernoulli random variables:

$$(B_n) : (\Omega, \mathcal{F}) \to (\{0, 1\}, 2^{\{0,1\}}),$$

and $\mathbf{B}$ is the random vector $(B_1, B_2, \ldots)$. Note that setting $Y = \cdot B_1 B_2 \ldots$ [it being understood that a dyadic representation is being given] it follows (from the fact that the Borel sets of $[0, 1]$ are generated by the intervals with dyadic rational endpoints) that Y is a random variable:

$$Y : (\Omega, \mathcal{F}) \to ([0, 1], \mathcal{B}([0, 1])).$$

Example: Skorokhod weak does not imply Skorokhod strong convergence. Essentially we just want an example of a sequence of densities which converge in probability, but not almost surely. Given the (B_n), define $\mathbb{P}^k$ as follows: express $k = 2^n + r$ $(0 \leq r \leq 2^n - 1)$, then

(i) under $\mathbb{P}^k$, $(B_1, \ldots, B_n, B_{n+2}, \ldots)$ are iid Bernoulli (parameter $\frac{1}{2}$);

(ii) if $\cdot B_1 \ldots B_n$ is **not** the dyadic representation of $\frac{r}{2^n}$ then make B_{n+1} conditionally independent Bernoulli $(\frac{1}{2})$;

(iii) if $\cdot B_1 \ldots B_n$ *is* the representation of $\frac{r}{2^n}$, then set $B_{n+1} = 1$.

It follows that f_k, the density of $\mathbb{P}^k_Y$ is given by

$$f_k(x) = \begin{cases} 1 : & x \notin [\frac{r}{2^n}, \frac{r+1}{2^n}) \\ 2 : & x \in [\frac{r+\frac{1}{2}}{2^n}, \frac{r+1}{2^n}) \\ 0 : & x \in [\frac{r}{2^n}, \frac{r+\frac{1}{2}}{2^n}) \end{cases}$$

where k (as before) is $2^n + r$ (with $0 \leq r \leq 2^n - 1$). Clearly $f_n \xrightarrow{prob} f_\infty (\equiv 1)$, since f_n differs from f_∞ only on a set of Lebesgue measure $O(\frac{1}{log_2 n})$, but equally clearly

$$\liminf f_n = 0 \text{ Lebesgue a.e.}$$

□

Example: Skorokhod strong convergence does not imply a.s convergence of densities. Here we just content ourselves with giving f_k:

$$f_k(x) = \begin{cases} 1 - 2^{-n} : & x \notin [\frac{r}{2^n}, \frac{r+1}{2^n}) \\ 2 - 2^{-n} : & x \in [\frac{r}{2^n}, \frac{r+1}{2^n}) \end{cases}$$

where, as usual, $k = 2^n + r$ $(0 \leq r \leq 2^n - 1)$. Clearly,

$$\liminf f_n = 1,$$

but

$$\limsup f_n = 2 \text{ (Lebesgue a.e.)}$$

□

An example from finance (see [7])

Suppose X^1 and X^2 are two continuous-time, skip-free Markov chains on $S = \{0, 1, 2, ..., d\}$ (generalised birth-and-death processes), with birth rates λ_n^i and death rates μ_n^i and suppose we know that $X_0^1 = x \geq X_0^2 = y$ and $\lambda_n^1 \geq \lambda_n^2$ and $\mu_n^1 \leq \mu_n^2$ for each n.

Q: How do we show that X_t^1 stochastically dominates X_t^2, i.e. that

$$\mathbb{P}(X_t^1 \geq k) \geq \mathbb{P}(X_t^2 \geq k)$$

for each k and t?

A: by a suitable coupling, using Poisson-thinning. Poisson thinning is, at its simplest, the act of obtaining a Poisson($(1-q)\theta$) process from Y, a Poisson(θ) process, by removing jumps of Y independently with probability q.

Let $\max_{n\in S}(\lambda_n^1+\mu_n^2)=\rho$. Construct a probability space with N, a Poisson process with rate 2ρ, and independent $U[0,1]$ r.v.s $U_1,\ldots$. Now start versions of X^i at x and y respectively and construct them as follows:

at T_k, the time of the kth jump of N, suppose $X^i_{T_k-}=x^i_k$, then whenever $x^2_k<x^1_k$, X^1 jumps down by 1 if $U<\frac{\mu^i_k}{2\rho}$, jumps up by 1 if $\frac{1}{2}>U>\frac{1}{2}-\frac{\lambda^i_k}{2\rho}$ otherwise X^1 doesn't move.

Similarly, X^2 jumps down by 1 if $\frac{1}{2}<U<\frac{1}{2}+\frac{\mu^2_k}{2\rho}$, jumps up by 1 if $U>1-\frac{\lambda^2_k}{2\rho}$ otherwise X^1 doesn't move. Note that X^1 and X^2 cannot jump at the same time in this case.

However, if $x_k^1 = x_k^2$ then X^i jumps down by 1 if $U < \frac{\mu_k^i}{2\rho}$, jumps up by 1 if $U > 1 - \frac{\lambda_k^i}{2\rho}$ otherwise X^i doesn't move. Note that in this case, if X^1 jumps down, then so must X^2, while if X^2 jumps up then so must X^1.

Since the resulting constructions cannot jump over one another, we see that $X^1 \geq X^2$.

Exercise:*[Ex3] Let N be a Poisson(θ) process and construct M a Poisson($p\theta$) process by thinning N. Find $\mathbb{P}$(N and M differ on* $[0, T]$*) when* $T = \theta = 2$ *and* $p = 0.95$*. Using your favourite calculation package, compare this to the total variation distance between the distributions of N and M on [0,T].*

Convergence of Markov chains [See [8]] We assume that $(\mathbb{P}^n)$ are a collection of probability measures on $D([0,\infty);\mathbb{Z}^+)$: under $\mathbb{P}^n$, X (given by $X_t(\omega)=\omega_t$) is a time-inhomogeneous *non-explosive* Markov chain with initial distribution (p_i^n). We assume the existence of a dominating measure μ (finite on compact sets) with respect to which each probability measure has transition rates $q_{i,j}^n(t)$ $(t\geq 0, i,j\in\mathbb{Z})$ and, as usual we write $q_i^n(t)=-q_{i,i}^n(t)$.

Now fix $T>0$ (temporarily) and denote the restriction of the $(\mathbb{P}^n)$ to the paths of X on $[0,T]$ by $\mathbb{P}^n|_{[0,T]}$.

Theorem:[MC] (a) If

$$p_i^n \to p_i^\infty \text{ as } n \to \infty \text{ for each } i; \tag{4}$$

$$q_i^n \xrightarrow{L^1(\mu)} q_i^\infty \text{ as } n \to \infty \text{ for each } i; \quad \textit{and} \tag{5}$$

$$q_{i,j}^n \xrightarrow{\mu a.e.} q_{i,j}^\infty \textit{ for each } i \textit{ and } j \textit{ in } \mathbb{Z}^+; \tag{6}$$

then $\mathbb{P}^n|_{[0,T]} \overset{Ss}{\Rightarrow} \mathbb{P}^\infty|_{[0,T]}$.
(b) If (4) and (5) hold and

$$q_{i,j}^n \xrightarrow{(\mu)} q_{i,j}^\infty \textit{ for each } i,j \textit{ in } \mathbb{Z}^+ \tag{7}$$

then for, each $T > 0$,

$$\mathbb{P}^n|_{[0,T]} \overset{Sw}{\Rightarrow} \mathbb{P}^\infty|_{[0,T]}$$

Remark: We stress that we are assuming that, under $\mathbb{P}^\infty$, the chain is non-explosive.

Proof: We give first a dominating (probability) measure $\mathbb{Q}$: it is specified by having waiting time distribution "exponential(μ)" in each state, i.e. $q_i(t) \equiv 1$ for each i. Under $\mathbb{Q}$, the jump chain forms a sequence of iid geometric($\frac{1}{2}$) r.v.s so that $q_{i,j}(t) = 2^{-(j+1)}$ and $\mathbb{Q}(X_0 = i) = 2^{-(i+1)}$. We assume that μ is continuous i.e. non-atomic. It is then fairly clear that the density of $\mathbb{P}^k|_{[0,T]}$ wrt $\mathbb{Q}|_{[0,T]}$ *is* $f^k \equiv f^k_T$ *given by*

$$f^k(\omega) = \quad p^k_{\omega_0} \prod_{n=1}^N 2^{(\omega_{T_n}+1)} q^k_{\omega_{T_{n-1}},\omega_{T_n}}(T_n) \\ \times \prod_{n=1}^N \exp\left(\int_{T_{n-1}}^{T_n} (1 - q^k_{\omega_{T_{n-1}}}(t)) d\mu(t)\right) \\ \times \exp\left(\int_{T_N}^{T} (1 - q^k_{\omega_{T_N}}(t)) d\mu(t)\right), \qquad (8)$$

where $N = N^T(\omega) = \# \{jumps\ of\ X\ on\ [0, T]\}$, $T_0 = 0$, and T_n $(1 \leq n \leq N)$ are the successive jump times of X (on $[0, T]$). Finally, since under $\mathbb{Q}$ the chain is non-explosive, notice that for any $\varepsilon > 0$, there is an $n(\varepsilon)$ s.t. $\mathbb{Q}(N > n) \leq \frac{\varepsilon}{2}$ and then $\exists$ $m(n(\varepsilon), \varepsilon)$ s.t.

$$\mathbb{Q}(X \text{ leaves } \{0, \ldots, m\} \text{ before } T) \leq \frac{\varepsilon}{2}.$$

Denote the union of the two sets involved in these statements by A_ε. We are now ready to prove (a).

Under the assumption (5)

$$e^{-\int_u^v q_i^k(t)d\mu(t)} \to e^{-\int_u^v q_i^\infty(t)d\mu(t)},$$

for any $0 \le u \le v \le T$. *Hence, off* A_ε, *there are only finitely many terms in (8) and (by (4) and (6)) each converges* $\mathbb{Q}$ *a.s. to the corresponding term in* f^∞. *Thus* $\mathbb{Q}(f^k \not\to f^\infty) \le \mathbb{Q}(A_\varepsilon) \le \varepsilon$ *and since* ε *is arbitrary we have established (a).*

To prove (b) we need only to use the subsequence characterisation of convergence in probability. Given a subsequence (n_k) *take a sub-subsequence* (n_{k_j}) *(by diagonalisation), along which (6) holds (at least for* $t \in [0, T]$*) then* $f^{n_{k_j}} \stackrel{\mathbb{Q}a.s.}{\longrightarrow} f^\infty$ *as* $j \to \infty$ *by (a). The subsequence is arbitrary so* $f^n \stackrel{prob(\mathbb{Q})}{\longrightarrow} f^\infty$ $\square$

Exercise:[Ex4] Check that $\mathbb{P}^k(A) = \int_A f^k d\mathbb{Q}$ for a suitable (characterising) family of events A.

Remark: The proof of Theorem [MC] only deals with the case where μ is non-atomic; if μ has atoms there is no great additional difficulty: we simply need to replace $\exp(-\int f d\mu)$ by $\exp(-\int f d\mu^c)\prod(1 - f\Delta\mu)$ wherever such terms appear in (8). This, in particular, allows us to deal with the discrete-time case.

Remark: It's easy to amend the proof to deal with semi-Markov processes.

Conditioning Markov chains Suppose that, under $\mathbb{P}_x$, X is a Markov chain started at x and τ is a hitting time for X i.e. the first time that X hits some set. Define $\mathbb{P}_x^T$ to be the law of X conditional on the event $(\tau > T)$. Define

$$h(x,t) = \mathbb{P}_x(\tau > t),$$

then for $A \in \mathcal{F}_t$,

$$\mathbb{P}_x^T(A) = \frac{\mathbb{P}_x(A \cap (\tau > T))}{\mathbb{P}_x(\tau > T)} = \frac{\mathbb{E}_x[1_A f(X_t, T-t)]}{f(x,T)}.$$

So, for any fixed $S < T$, on $\mathcal{F}_S$

$$\frac{d\mathbb{P}_x^T}{d\mathbb{P}_x} = 1_{(\tau > S)} \frac{f(X_S, T-S)}{f(x,T)}$$

so if $\frac{f(y,T-S)}{f(x,T)} \to \rho_{x,S}(y)$ for each y where $1_{(\tau>S)}\rho$ is a density wrt $\mathbb{P}_x$ then $\mathbb{P}_x^T \overset{S_S}{\Rightarrow} \mathbb{P}_x^\infty$.

Example: *Suppose that, under $\mathbb{P}_x$, X is a MC on $\{0,1,2,\dots,d\}$ with Q-matrix Q and τ is the hitting time of 0. The trick here is to look at the asymptotic behaviour of $\tilde{\mathbb{P}}(t)$, the defective transition probabilities for the process killed on hitting 0. So look at $\tilde{Q}$, the restriction of Q to $\{1,2,\dots,d\}$, then $\tilde{\mathbb{P}}_t$ is*

$$\tilde{\mathbb{P}} = \exp(t\tilde{Q}).$$

Now if we can diagonalise $\tilde{Q}$ as $\tilde{Q} = E\Lambda D$ then, assuming we have written the largest eigenvalue of $\tilde{Q}$ as $\lambda = \Lambda_{1,1}$, $\tilde{\mathbb{P}}_t \sim e^{-\lambda t} e_{i,1} d_{1,j}$ and it follows that $f(i,t) \sim \sum_j e^{-\lambda t} e_{i,1} d_{1,j}$ and thus $\rho_{i,S}(j) = e^{\lambda S} h(j)/h(i)$, where $h = e_{\cdot,1}$ is the right eigenvector of $\tilde{Q}$ corresponding to the principal eigenvalue. Since h is *an eigenvector it follows that ρ is a density and so $\mathbb{P}^T \overset{Ss}{\Rightarrow} \mathbb{P}^\infty$ on $[0,S]$, where $\mathbb{P}^\infty$ is the law of a MC on with Q-matrix $\bar{Q}$ given by $\bar{Q}_{i,j} = \lambda\delta_{i,j} + h_j\tilde{Q}_{i,j}/h(i)$.*

Theorem:[Girsanov] Suppose that, under $\mathbb{P}$, B is a standard Brownian Motion, μ is a bounded continuous adapted process and we define $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(\int_0^T \mu_s dB_s - \frac{1}{2}\int_0^T \mu_s^2 ds\right),$$

then under $\mathbb{Q}$, $Z_t \stackrel{def}{=} B_t - \int_0^t \mu_s ds$ is a standard BM, or B is a BM with drift rate μ.

This allows us to copy what we did for convergence of Markov chains in the Ito diffusion setting.

Feller property The idea is to couple two copies X^x and X^y of a diffusion starting at different positions by coupling the driving BMs using a mirror coupling. This gives a lower bound on the probability of coupling by time T (note, they don't actually have to have the same SDE but...). So

$$X_t^z = z + \int_0^t \sigma(X_s^z)dB_s + \int_0^t \mu(X_s^z)ds, \tag{9}$$

where B is a standard BM and σ is bounded below by $\eta > 0$ and $|\mu| \leq M$. Take a copy of X^y, call it Y, where the SDE (9) is driven by $-B$ (also a BM). This will have the same law. Now take another copy, call it Z, driven by $-B$ until the stopping time τ and then driven by B, where τ is the coupling time:

$$\tau = \inf\{t : X_t = Y_t\}.$$

Now suppose that g is a bounded measurable function,then

$$|\mathbb{E}(g(X_T) - \mathbb{E}(g(Z_T)| \leq 2M\mathbb{P}(\tau > T)$$

where $M = \sup|g|$. Wlog $x > y$, so

$$\mathbb{P}(\tau > T) =$$

$$\mathbb{P}(x-y+\inf_{0\leq t\leq T}\big[\int_0^t(\sigma(X_s)+\sigma(Y_s))dB_s+\int_0^t(\mu(X_s)-\mu(Y_s))ds\big] > 0)$$
$$\leq \mathbb{P}_{x-y}(\tau_B > 4\eta^2 T),$$

where B is a one-dimensional BM with drift $\frac{M}{2\eta^2}$ and τ_B is the first time that B hits the origin. Now $\mathbb{P}_z(\tau_B > T)$ is $O(\frac{z}{\sqrt{T}})$ (as $z \to 0$) and hence we obtain the required result (see ([6]) and ([10])).

[1] R. Atar and K. Burdzy, Mirror Couplings and Neumann Eigenfunctions. *Journal of the American Mathematical Society* 17(2), 243–265, 2004.

[2] M. T. Barlow and S. D. Jacka, Tracking a Diffusion, and an Application to Weak Convergence. *Advances in Applied Probability* 18, 15–25, 1986.

[3] K. Burdzy and W. S. Kendall, Efficient Markovian Couplings: Examples and Counterexamples. *The Annals of Probability* 10(2), 362–409, 2000.

[4] E. P. Hsu and K.T. Sturm, Maximal Coupling of Euclidean Brownian Motions. SFB Preprint 85, University of Bonn, 2003.

[5] Saul Jacka and Aleksandar Mijatović, Coupling and Tracking of Regime-Switching Martingales. ArXiv:1209.0180, 2012.

[6] Saul Jacka, Aleksandar Mijatovic and Dejan Siraj, Mirror and synchronous couplings of Geometric Brownian motion, Stochastic Processes and their Applications, 124, 1055-1069 (2014)

[7] Saul Jacka and Adriana Ocejo, American-type options with parameter uncertainty, arXiv:1309.1404 , 2013.

[8] Saul Jacka and Gareth Roberts, 'On strong forms of weak convergence' Stoch. Proc. & Appl. 67, 41-53 (1997)

[9] T. Lindvall, *Lectures on the Coupling Method*. Dover Publications, New York, 2002.

[10] T. Lindvall and L. C. G. Rogers, Coupling of Multidimensional Diffusions by Reflection. *The Annals of Probability* 14(3), 860–872, 1986.

[11] M. N. Pascu, Mirror Coupling of Reflecting Brownian motion and an Application to Chavel's Conjecture. *Electronic Journal of Probability* 16, 505–530, 2011.

[12] H. Thorisson, *Coupling, Stationarity, and Regeneration*. Springer-Verlag New York, 2000.