On Coupling and Convergence in Density and in Distribution

Hermann Thorisson

Abstract

According to Dudley's extension of the Skorohod representation theorem, convergence in distribution on a separable metric space is equivalent to the existence of a coupling with elements converging a.s. in the metric. A density analogue of this theorem says that a sequence of probability densities on a general measurable space has a probability density as a lower pointwise limit if and only if there exists a coupling with elements converging a.s. in the discrete topology. In this paper the latter result is extended to discrete-topology convergence of stochastic processes in a widening time-window. An elementary version of that result is then used to prove the Skorohod-Dudley theorem.

1 Introduction and statement of results

The aim of this paper is to present a coupling characterization of finite-windows density convergence of a sequence of stochastic processes (Theorem 1), and to use an elementary version of that result (Corollary 1) to prove Dudley's extension of the Skorohod representation theorem. Recall that with I some index set, a *coupling* of a collection of random elements X_i, $i \in I$, is a family $(\hat{X}_i : i \in I)$ such that for each $i \in I$, $\hat{X}_i$ has the same distribution as X_i. For convenience, let all random elements in this paper be defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

The following celebrated theorem was proved by Skorohod (1956) in the Polish (i.e. complete separable) case, and extended to the separable case by Dudley (1968). For historical notes, see Dudley (2002).

MSC 2000 *subject classifications.* Primary 60F15; Secondary 60G99.
Key words and phrases. Skorohod representation, weak convergence, density convergence.
Running head. Coupling and Convergence in Density and in Distribution

Skorohod-Dudley Theorem. *Let $X_1, X_2, \ldots, X$ be random elements in a separable metric space E endowed with its Borel subsets $\mathcal{E}$. Then*

$$X_n \to X \text{ in distribution with respect to the metric, as } n \to \infty, \tag{1}$$

if and only if there exists a coupling $(\hat{X}_1, \hat{X}_2, ..., \hat{X})$ of $X_1, X_2, ..., X$ such that

$$\hat{X}_n \to \hat{X} \text{ a.s. in the metric, as } n \to \infty.$$

While this theorem does not have a simple proof, the following density analogue from Thorisson (1995) is easy to establish. Note that random elements $X_1, X_2, \ldots$ in an arbitrary space always have densities with respect to some measure λ, for instance with respect to $\lambda = \sum_{n=1}^{\infty} 2^{-n}\mathbf{P}(X_n \in \cdot)$.

Proposition 1. *Let $X_1, X_2, \ldots, X$ be random elements in an arbitrary measurable space $(E, \mathcal{E})$. Let $f_1, f_2, \ldots$ be the densities of $X_1, X_2, \ldots$ with respect to some measure λ. Then*

$$\liminf_{n\to\infty} f_n \text{ is a density of } X \text{ with respect to } \lambda \tag{2}$$

if and only if there exists a coupling $(\hat{X}_1, \hat{X}_2, \ldots, \hat{X})$ of $X_1, X_2, \ldots, X$ and an $\mathbb{N}$-valued random variable N such that

$$\hat{X}_n = \hat{X}, \quad n \geqslant N. \tag{3}$$

Proof. Assume existence of the coupling. Take $n < m$ and partition E into sets $A_n, \ldots, A_m \in \mathcal{E}$ such that $\min_{n\leqslant i\leqslant m} f_i = f_j$ on A_j for $n \leqslant j \leqslant m$. This yields the last step in the following calculation: for $A \in \mathcal{E}$,

$$\begin{aligned}\mathbf{P}(\hat{X} \in A, N \leqslant n) &= \sum_{j=n}^{m} \mathbf{P}(\hat{X}_j \in A \cap A_j, N \leqslant n) \quad \text{[partition and (3)]}\\ &\leqslant \sum_{j=n}^{m} \mathbf{P}(\hat{X}_j \in A \cap A_j) = \sum_{j=n}^{m} \int_{A\cap A_j} f_j \, d\lambda = \int_A \min_{n\leqslant i\leqslant m} f_i \, d\lambda.\end{aligned}$$

Send first m and then n to infinity to obtain $\mathbf{P}(X \in A) \leqslant \int_A \liminf_{n\to\infty} f_n \, d\lambda$ for all $A \in \mathcal{E}$. This forces the inequality to be an identity for all $A \in \mathcal{E}$.

Conversely, assume that $g_n := \inf_{m\geqslant n} f_m$ increases to a density of X as $n \to \infty$. Let N have distribution function $\mathbf{P}(N \leqslant n) = \int g_n d\lambda$, $n \in \mathbb{N}$. Let $V_1, V_2, \ldots, W_1, W_2, \ldots$ be independent random elements in $(E, \mathcal{E})$ that are independent of N. For $n \in \mathbb{N}$, let V_n have density $(g_n - g_{n-1})/\mathbf{P}(N = n)$, let W_n have density $(f_n - g_n)/\mathbf{P}(N > n)$, and put $\hat{X}_n = V_N$ on $\{N \leqslant n\}$ and $\hat{X}_n = W_n$ on $\{N > n\}$. Put $\hat{X} = V_N$ to obtain the coupling result. $\square$

The following theorem extends this result on convergence in the discrete topology to convergence of stochastic processes *in a widening time-window.* We use the notation $\mathbf{X} = (X^s)_{s\in[0,\infty)}$ for a continuous-time stochastic process, and $\mathbf{X}^t = (X^s)_{s\in[0,t)}$ for a finite segment of the process of length $t \in [0,\infty)$.

Theorem 1. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}$ be continuous-time stochastic processes on a Polish state space $(E, \mathcal{E})$ with right-continous paths having left-hand limits. For each $t \in [0,\infty)$, let $f_1^t, f_2^t, \dots$ be the densities of $\mathbf{X}_1^t, \mathbf{X}_2^t, \dots$ with respect to some measure λ^t. Then*

$$\forall t \in [0,\infty): \quad \liminf_{n\to\infty} f_n^t \text{ is a density of } \mathbf{X}^t \text{ with respect to } \lambda^t \tag{4}$$

if and only if there exist non-negative numbers $t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n \to \infty$ as $n \to \infty$, a coupling $(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \dots, \hat{\mathbf{X}})$ of $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}$, and an $\mathbb{N}$-valued random variable N such that

$$\hat{\mathbf{X}}_n^{t_n} = \hat{\mathbf{X}}^{t_n}, \quad n \geqslant N.$$

It follows as a corollary, or by repeating the proof with the appropriate modifications, that the same holds for discrete-time stochastic processes on a Polish state space. In particular, Theorem 2 has the following corollary, where we use the notation $\mathbf{X} = (X^1, X^2, \dots)$ for a discrete-time stochastic process and $\mathbf{X}^k := (X^1, \dots, X^k)$ for a segment of integer length $k \geqslant 0$.

Corollary 1. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}$ be discrete-time stochastic processes on a countable state space E. If*

$$\forall k \in \mathbb{N}: \quad \mathbf{P}(\mathbf{X}_n^k = \mathbf{x}^k) \to \mathbf{P}(\mathbf{X}^k = \mathbf{x}^k) \text{ as } n \to \infty, \ \ \mathbf{x}^k \in E^k, \tag{5}$$

then there exist non-negative integers $k_1 \leqslant k_2 \leqslant \cdots \leqslant k_n \to \infty$ as $n \to \infty$, a coupling $(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \dots, \hat{\mathbf{X}})$ of $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}$, and an $\mathbb{N}$-valued random variable N such that

$$\hat{\mathbf{X}}_n^{k_n} = \hat{\mathbf{X}}^{k_n}, \quad n \geqslant N.$$

It is readily checked that the implication in the corollary can be reversed.

We prove Theorem 1 in Section 2, and use the corollary in Section 3 to prove the Skorohod-Dudley Theorem. In fact, Corollary 1 is a relatively elementary result and we include a direct proof in Section 4.

2 Proof of Theorem 1

Assume existence of the coupling. For $t \in [0,\infty)$, take $m \in \mathbb{N}$ such that $t_m \geqslant t$ and note that $\hat{\mathbf{X}}_n^t = \hat{\mathbf{X}}^t$, $n \geqslant \max\{N, m\}$. Apply Proposition 1 to obtain (4).

Conversely, assume (4). Regard the processes as random elements in Skorohod space, denoted $(D, \mathcal{D})$ and $(D^t, \mathcal{D}^t)$ for the time sets $[0, \infty)$ and $[0, t)$, respectively. Let $P_1, P_2, \ldots, P$ be the distributions of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}$, and $P_1^t, P_2^t, \ldots, P^t$ the distributions of $\mathbf{X}_1^t, \mathbf{X}_2^t, \ldots, \mathbf{X}^t$. Note that $P_1^0, P_2^0, \ldots, P^0$ all have mass one at the empty vector. Let ν_n^t be the greatest common component of the measures $P_n^t, P_{n+1}^t, \ldots$, that is, let ν_n^t be the measure with density $\inf_{m \geqslant n} f_m^t$ with respect to λ^t. Let $\|\cdot\|$ denote mass. Use (4) to find non-negative numbers $t_1 \leqslant t_2 \leqslant \cdots \to \infty$ such that $\|P - \nu_n^{t_n}\| \leqslant 2^{-n}$, $n \in \mathbb{N}$. Due to (4), for all $n \in \mathbb{N}$ we have $\nu_n^{t_n} \leqslant P^{t_n}$. Use this to extend $\nu_n^{t_n}$ from $(D^{t_n}, \mathcal{D}^{t_n})$ to a measure ν_n on $(D, \mathcal{D})$ by $\nu_n(A) := \int \mathbf{P}(X \in A | X^{t_n} = \cdot) d\nu_n^{t_n}$, $A \in \mathcal{D}$. Then $\nu_n \leqslant P$. Let μ_n be the greatest common component of ν_m, $m \geqslant n$. Let $A_m \in \mathcal{D}$, $m \geqslant n$, be a partition of D such that $\mu_n(\cdot \cap A_m) = \nu_m(\cdot \cap A_m)$. Then $P - \mu_n = \sum_n^\infty (P(\cdot \cap A_m) - \nu_m(\cdot \cap A_m)) \leqslant \sum_n^\infty 2^{-m} = 2^{n+1} \to 0$ as $n \to \infty$. Thus μ_n increases setwise to P as $n \to \infty$. Let μ_n^t be the restriction of μ_n to $(D^t, \mathcal{D}^t)$ and note that $\mu_n^{t_n} \leqslant \nu_n^{t_n}$. Thus $\mu_n^{t_n} \leqslant P_n^{t_n}$, $n \in \mathbb{N}$. Put $\mu_0 := 0$.

Now, let $N, \mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{W}_1, \mathbf{W}_2, \ldots$ be independent. Let N have distribution function $\mathbf{P}(N \leqslant n) = \|\mu_n\|$, $n \in \mathbb{N}$. For $n \in \mathbb{N}$, let $\mathbf{V}_n$ be a random element in $(D, \mathcal{D})$ with distribution $(\mu_n - \mu_{n-1})/\mathbf{P}(N = n)$, let $\mathbf{W}_n$ be a random element in $(D^{t_n}, \mathcal{D}^{t_n})$ with distribution $(P_n^{t_n} - \mu_n^{t_n})/\mathbf{P}(N > n)$, and define $\hat{\mathbf{X}}_n^{t_n} = \mathbf{V}_N^{t_n}$ on $\{N \leqslant n\}$ and $\hat{\mathbf{X}}_n^{t_n} = \mathbf{W}_n$ on $\{N > n\}$. Then $\hat{\mathbf{X}}_n^{t_n}$ has distribution $P_n^{t_n}$, which is the distribution of $\mathbf{X}_n^{t_n}$. Use this, the Ionescu-Tulcea Extension Theorem, and the existence [since $(D, \mathcal{D})$ is Polish] of a regular version of the conditional distribution of $(X_n^{t_n+s})_{s \in [0,\infty)}$ given the value of $\mathbf{X}_n^{t_n}$, to extend each $\hat{\mathbf{X}}_n^{t_n}$ to a full process $\hat{\mathbf{X}}_n$ with the same distribution as $\mathbf{X}_n$. Finally, note that $\hat{\mathbf{X}} := \mathbf{V}_N$ has the same distribution as $\mathbf{X}$.

3 Proof of the Skorohod-Dudley Theorem

Assume existence of the coupling. Let h be bounded and continuous to obtain that $h(\hat{X}_n) \to h(\hat{X})$ a.s. as $n \to \infty$. Then by bounded convergence $\mathbf{E}[h(X_n)] \to \mathbf{E}[h(X)]$ as $n \to \infty$. Thus by definition, (1) holds.

Conversely, assume (1). Let d be the metric and put $P := \mathbf{P}(X \in \cdot)$. Recall that $A \in \mathcal{E}$ is a *P-continuity set* if $P(\partial A) = 0$ where ∂A denotes the boundary of A, and that for such A the Portmanteau Theorem [Theorem 11.1.1 in Dudley (2002)] implies that $\mathbf{P}(X_n \in A) \to P(A)$ as $n \to \infty$. By separability, for each $\epsilon > 0$ the set E can be covered by countably many balls of diameter $< \epsilon$. Note that $\partial\{x \in E : d(y,x) < r\} \subseteq \partial\{x \in E : d(y,x) = r\}$ and that the sets on the right have P-mass 0 except for countably many radii r. Thus the covering sets may be taken to be P-continuity sets. Moreover, since $\partial(A \cap B) \subseteq \partial A \cup \partial B$ the covering sets can be taken to be disjoint.

Let $\{A_1, A_2, \dots\}$ be a partition of E into P-continuity sets of diameter < 1. For $i \in \mathbb{N}$, let $\{A_{i1}, A_{i2}, \dots\}$ be a partition of A_i into P-continuity sets of diameter $<1/2$. Continue recursively to obtain nested partitions $\{A_{\mathbf{i}^k} : \mathbf{i}^k \in \mathbb{N}^k\}$ of E into P-continuity sets of diameter $< 1/k$, $k \in \mathbb{N}$.

After these standard preliminaries, we are now ready to apply first Portmanteau and then Corollary 1: with $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}$ the discrete-time processes on $\mathbb{N}$ defined by (well-defined because the partitions are nested)

$$\mathbf{M}_n^k := \mathbf{i}^k \text{ if } X_n \in A_{\mathbf{i}^k} \qquad \text{and} \qquad \mathbf{M}^k := \mathbf{i}^k \text{ if } X \in A_{\mathbf{i}^k},$$

we have $\mathbf{P}(\mathbf{M}_n^k = \mathbf{i}^k) \to \mathbf{P}(\mathbf{M}^k = \mathbf{i}^k)$ as $n \to \infty$, and thus there are non-negative integers $k_1 \leqslant k_2 \leqslant \dots \leqslant k_n \to \infty$ as $n \to \infty$, a coupling $(\hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2, \dots, \hat{\mathbf{M}})$ of $\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}$, and an $\mathbb{N}$-valued N such that

$$\hat{\mathbf{M}}_n^{k_n} = \hat{\mathbf{M}}^{k_n}, \quad n \geqslant N. \tag{6}$$

Since the family $(N, \hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2, \dots)$ consists of countably many discrete random variables, there exists a regular version of its conditional distribution given the value of $\hat{\mathbf{M}}$. Thus by extending the underlying probability space, we can take $(N, \hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2, \dots, \hat{\mathbf{M}})$ such that $\hat{\mathbf{M}} = \mathbf{M}$. Put $\hat{X} := X$.

Let $(V_{n,\mathbf{i}^{k_n}} : n \in \mathbb{N}, \mathbf{i}^{k_n} \in \mathbb{N}^{k_n})$ be a family of independent random elements. Let the family be independent of $(N, \hat{\mathbf{M}}_1, \hat{\mathbf{M}}_2, \dots, \hat{\mathbf{M}}, \hat{X})$. Let $V_{n,\mathbf{i}^{k_n}}$ be $A_{\mathbf{i}^{k_n}}$-valued with distribution $\mathbf{P}(X_n \in \cdot \mid X_n \in A_{\mathbf{i}^{k_n}})$. Put $\hat{X}_n := V_{n,\hat{\mathbf{M}}_n^{k_n}}$. Then

$$\mathbf{P}(\hat{X}_n \in \cdot) = \sum_{\mathbf{i}^{k_n} \in \mathbb{N}^{k_n}} \mathbf{P}(V_{n,\mathbf{i}^{k_n}} \in \cdot) \mathbf{P}(\hat{\mathbf{M}}_n^{k_n} = \mathbf{i}^{k_n}) = \mathbf{P}(X_n \in \cdot), \quad n \in \mathbb{N}.$$

By (6), $n \geqslant N$ implies $\hat{X}_n = V_{n,\hat{\mathbf{M}}^{k_n}}$ and thus $\hat{X}_n \in A_{\hat{\mathbf{M}}^{k_n}}$. But $\hat{X} \in A_{\hat{\mathbf{M}}^{k_n}}$ for all $n \in \mathbb{N}$. Thus $n \geqslant N$ implies $d(\hat{X}_n, \hat{X}) \leqslant 1/k_n$, which goes to 0 as $n \to \infty$.

4 A more elementary proof of Corollary 1

Assume (5). The main part of the proof is a stepwise construction of a coupling $(\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \dots, \hat{\mathbf{X}})$ and $\mathbb{N}$-valued $N^1 \leqslant N^2 \leqslant \dots$ such that for $k \in \mathbb{N}$,

$$\hat{\mathbf{X}}_n^k = \hat{\mathbf{X}}^k, \quad n \geqslant N^k. \tag{7}$$

First note that (5) with $k = 1$ yields (2) with $f_n = \mathbf{P}(X_n^1 = \cdot)$ and λ counting measure. Thus the only-if part of Proposition 1 gives us the existence of a family $(\hat{X}_1^1, \hat{X}_2^1, \dots, \hat{X}^1, N^1)$ such that $(\hat{X}_1^1, \hat{X}_2^1, \dots, \hat{X}^1)$ is a coupling of $(X_1^1, X_2^1, \dots, X^1)$ and $\hat{X}_n^1 = \hat{X}^1$ for $n \geqslant N^1$. Thus (7) holds for $k = 1$.

Then, note that there are countably many $\mathbf{y^k} := (\mathbf{x}_1^k, \mathbf{x}_2^k, \dots, \mathbf{x}^k, m)$ such that $k, m \in \mathbb{N}$, $\mathbf{x}_1^k, \mathbf{x}_2^k, \dots, \mathbf{x}^k \in E^k$, $\mathbf{x}_n^k = \mathbf{x}^k$ for $n \geqslant m$ and $\mathbf{P}(\mathbf{X}^k = \mathbf{x}^k) > 0$, and that for each such $\mathbf{y^k}$ we obtain from (5) that

$$\mathbf{P}(X_n^{k+1} = \cdot \mid \mathbf{X}_n^k = \mathbf{x}_n^k) \to \mathbf{P}(X^{k+1} = \cdot \mid \mathbf{X}^k = \mathbf{x}^k), \quad n \to \infty.$$

Thus Proposition 1 gives us a family $(X_1^{\mathbf{y}^k}, X_2^{\mathbf{y}^k}, \dots, X^{\mathbf{y}^k}, N^{\mathbf{y}^k})$ such that

$$\begin{aligned}
&\mathbf{P}(X_n^{\mathbf{y}^k} = \cdot) = \mathbf{P}(X_n^{k+1} = \cdot \mid \mathbf{X}_n^k = \mathbf{x}_n^k), \quad n \geqslant 1, \\
&\mathbf{P}(X^{\mathbf{y}^k} = \cdot) = \mathbf{P}(X^{k+1} = \cdot \mid \mathbf{X}^k = \mathbf{x}^k), \\
&X_n^{\mathbf{y}^k} = X^{\mathbf{y}^k}, \quad n \geqslant N^{\mathbf{y}^k}.
\end{aligned} \tag{8}$$

Let these families be independent. Define the full coupling $(\hat{\mathbf{X}}_\mathbf{1}, \hat{\mathbf{X}}_\mathbf{2}, \dots, \hat{\mathbf{X}})$ recursively in $k \in \mathbb{N}$ by $(\hat{X}_1^{k+1}, \hat{X}_2^{k+1}, \dots, \hat{X}^{k+1}) := (X_1^{\mathbf{Y}^k}, X_2^{\mathbf{Y}^k}, \dots, X^{\mathbf{Y}^k})$ where $\mathbf{Y}^k := (\mathbf{X}_1^k, \mathbf{X}_2^k, \dots, \mathbf{X}^k, N^k)$. Put $N^{k+1} := \max\{N^k, N^{\mathbf{Y}^k}\}$ for $k \in \mathbb{N}$.

We have already established (7) for $k = 1$. Make the induction assumption that (7) holds for a $k \in \mathbb{N}$. By (8) and the definition of N^{k+1}, $\hat{X}_n^{k+1} = \hat{X}^{k+1}$ for $n \geqslant N^{k+1}$. This and the induction assumption imply that (7) also holds with k replaced by $k+1$. Thus by induction, (7) holds for all $k \in \mathbb{N}$.

In order to complete the proof of Corollary 1, let $n^1 < n^2 < \cdots < n^k \to \infty$ as $k \to \infty$ be positive integers such that $\mathbf{P}(N^k > n^k) \leqslant 1/k^2$, $k \in \mathbb{N}$. Put $K := \sup\{k \in \mathbb{N} : N^k > n^k\}$ and note that $K < \infty$ a.s. due to Borel-Cantelli. Let $n^0 = 0$ and for $n \in \mathbb{N}$ put $k_n = k$ if $n^k \leqslant n < n^{k+1}$. Define $N := n^{K+1}$. Take $n \geqslant N$ and note that then $k_n > K$ which implies that $n^{k_n} \geqslant N^{k_n}$. Now $n \geqslant n^{k_n}$ and thus $n \geqslant N^{k_n}$, and (7) yields $\hat{\mathbf{X}}_n^{k_n} = \hat{\mathbf{X}}^{k_n}$.

References

[1] Dudley, R.M. (1968). Distances of probability measures and random variables. *Annals of Mathematical Statistics* **39**, 1563–1572.

[2] Dudley, R.M. (2002). *Real Analysis and Probability.* Cambridge University Press, Cambridge.

[3] Skorohod, A.V. (1956). Limit theorems for stochastic processes. *Theo. Probab. Appl.* **1**, 261-290.

[4] Thorisson, H. (1995). Coupling methods in probability theory. *Scand. J. Statist.* **22**, 159–182.

Hermann Thorisson, Science Institute, University of Iceland,
Dunhaga 3, 107 Reykjavik, Iceland, hermann@hi.is