UDC 519.2

On Coupling and Convergence in Density and in Distribution

H. Thorisson*

* *University of Iceland, Dunhaga 3, 107 Reykjavik, Iceland*

Abstract. According to the Skorohod representation theorem, convergence in distribution to a limit in a separable set 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 pointwise lower limit if and only if there exists a coupling with elements converging a.s. in the discrete metric. In this note the discrete-metric theorem is extended to a sequence of stochastic processes considered in a widening time window. The extension is then used to prove the Skorohod representation theorem.

Keywords: coupling, widening time window, Skorohod representation.

1. Introduction

Let $X_1, X_2, \dots, X$ be random elements in some measurable space $(E, \mathcal{E})$ with distributions $P_1, P_2, \dots, P$, respectively. Let $(\hat{X}_1, \hat{X}_2, \dots, \hat{X})$ denote a coupling of $X_1, X_2, \dots, X$. This means that the random elements $\hat{X}_1, \hat{X}_2, \dots, \hat{X}$ are all defined on the same probability space and have the distributions $P_1, P_2, \dots, P$, respectively.

If E is a metric space and $\mathcal{E}$ its Borel subsets, write

$$X_n \to X \quad \text{in distribution as} \quad n \to \infty$$

to denote that for each bounded continuous function $h : E \to \mathbb{R}$

$$\int h \, dP_n \to \int h \, dP \quad \text{as} \quad n \to \infty.$$

The following result is the celebrated Skorohod Representation Theorem. It was proved by Skorohod in 1956 for random elements in a complete separable space, [2]. Dudley [1] removed the completeness assumption in 1968 and Wichura [5] showed in 1970 that it suffices that the support of P is separable.

Theorem 1. *Let E be a metric space, let $\mathcal{E}$ be its Borel subsets, and let P have a separable support. Then*

$$X_n \to X \quad \textit{in distribution as} \quad n \to \infty$$

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

$$\hat{X}_n \to \hat{X} \quad \textit{pointwise in the metric as} \quad n \to \infty. \qquad \triangleright$$

Here we shall present a proof of this theorem based on the concept of *convergence in density.* In Section 2, we introduce that concept and prove the associated representation (coupling) theorem which has a relatively simple proof. In Section 3, we extend that representation theorem to stochastic processes considered in a widening time window. In Section 4, we use the extension to prove the Skorohod Representation Theorem.

2. Convergence in Density

Let $f_1, f_2, \ldots, f$ be densities of $X_1, X_2, \ldots, X$ with respect to a measure λ. Note that such a λ always exists: we can for instance take $\lambda = P + \sum \frac{1}{2^i} P_i$. Let

$$X_n \to X \quad \text{in density as} \quad n \to \infty$$

denote that

$$\liminf_{n\to\infty} f_n = f \quad \text{a.e. } \lambda \text{ as } n \to \infty.$$

Note that f_n/f is defined a.e. with respect to P. It is the Radon-Nikodym density dP_n/dP of the absolutely continuous part of P_n with respect to P. Thus convergence in density is equivalent to

$$\liminf_{n\to\infty} dP_n/dP = 1 \quad \text{a.e. } P \quad \text{as } n \to \infty.$$

Note also that in the discrete case convergence in density means simply that $\lim_{n\to\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x), \ x \in E$.

In a 1995 paper [3], Section 5.4, this author established the following analog of the Skorohod Representation Theorem.

Theorem 2. *Let $(E, \mathcal{E})$ be a general measurable space. Then*

$$X_n \to X \quad \textit{in density as} \quad n \to \infty$$

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

$$\hat{X}_n \to \hat{X} \quad \textit{pointwise in the discrete metric as} \quad n \to \infty. \qquad \triangleright$$

Proof. First, assume the coupling claim, namely that there is an $\mathbb{N}$-valued random variable N such that $\hat{X}_n = \hat{X}, \ n \geq N$. Fix $B \in \mathcal{E}$ and $n < m$. Partition E into sets $A_n, \ldots, A_m \in \mathcal{E}$ such that $\min_{n \leq i \leq m} f_i = f_j$ on A_j for $n \leq j \leq m$. Then

$$\begin{aligned}
\mathbb{P}(\hat{X} \in B, N \leq n) &= \sum_{j=n}^{m} \mathbb{P}(\hat{X} \in B \cap A_j, N \leq n) \qquad \text{[partition]} \\
&= \sum_{j=n}^{m} \mathbb{P}(\hat{X}_j \in B \cap A_j, N \leq n) \qquad [\text{since } \hat{X}_j = \hat{X} \text{ when } j \geq n \geq N] \\
&\leq \sum_{j=n}^{m} \mathbb{P}(\hat{X}_j \in B \cap A_j) = \sum_{j=n}^{m} \int_{B\cap A_j} f_j = \int_B \min_{n \leq i \leq m} f_i \leq 1\,.
\end{aligned}$$

Send first $m \to \infty$ to obtain that $\mathbb{P}(\hat{X} \in B, N \leq n) \leq \int_B \inf_{n \leq i < \infty} f_i \leq 1$. Then send $n \to \infty$ to obtain that

$$P(B) \;\leq\; \int_B \liminf_{n\to\infty} f_n \;\leq 1 \qquad \text{for all } B \in \mathcal{E}.$$

This forces $P(B) = \int_B \liminf_{n\to\infty} f_n$ for all $B \in \mathcal{E}$. Thus the if-direction holds.

Conversely, assume that $\liminf_{n\to\infty} f_n$ is a density of X. Let μ_n be the measure with density $g_n := \inf_{m \geq n} f_m$ w.r.t. λ. Then $\mu_1 \leq \mu_2 \leq \cdots \nearrow P$. Let $N, V_1, V_2, \ldots, W_1, W_2, \ldots$ be independent random elements such that

N has distribution function $\quad \mathbb{P}(N \leq n) = \mu_n(E), \quad n \in \mathbb{N},$

V_n has distribution $\;(\mu_n - \mu_{n-1})/\mathbb{P}(N = n)$ where $\;\mu_0 = 0,$

W_n has distribution $\;(P_n - \mu_n)/\mathbb{P}(N > n).$

Put $\hat{X}_n = V_N$ on $\{N \leq n\}$ and $\hat{X}_n = W_n$ on $\{N > n\}$. Then

$$\begin{aligned}\mathbb{P}(\hat{X}_n \in \cdot) \;&=\; \sum_{1 \leq i \leq n} \mathbb{P}(V_i \in \cdot)\mathbb{P}(N = i) + \mathbb{P}(W_n \in \cdot)\mathbb{P}(N > n) \\ &=\; \sum_{1 \leq i \leq n} (\mu_i - \mu_{i-1}) + (P_n - \mu_n) \;=\; P_n\,.\end{aligned}$$

Put $\hat{X} = V_N$. Then (again by telescoping)

$$\mathbb{P}(\hat{X} \in \cdot) = \sum_{1 \leq i < \infty} \mathbb{P}(V_i \in \cdot)\mathbb{P}(N = i) = \sum_{1 \leq i < \infty} (\mu_i - \mu_{i-1}) \;=\; P.$$

Clearly $\hat{X}_n = \hat{X},\; n \geq N$. Thus the only-if-direction also holds. □

3. Convergence in Density – Widening Time Window

Set $(H, \mathcal{H}) = (E^{(1)}, \mathcal{E}^{(1)}) \otimes (E^{(2)}, \mathcal{E}^{(2)}) \otimes \cdots \otimes (E, \mathcal{E})$. If $\mathbf{Z} = (Z^1, Z^2, \ldots, X)$ is a random element in $(H, \mathcal{H})$ write $\mathbf{Z}^k = (Z^1, Z^2, \ldots, Z^k)$ for a time window of length $k \in \mathbb{N}$. Set $(H^k, \mathcal{H}^k) = (E^{(1)}, \mathcal{E}^{(1)}) \otimes \cdots \otimes (E^{(k)}, \mathcal{E}^{(k)})$.

The following theorem is from the author's 2016 paper [4].

Theorem 3. *Let* $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}$ *be random elements in* $(H, \mathcal{H})$*. Then*

$$\forall k \in \mathbb{N}: \quad \mathbf{Z}_n^k \to \mathbf{Z}^k \;\; \textit{in density as} \;\; n \to \infty$$

if and only if there are integers $0 \leq k_1 \leq k_2 \leq \cdots \to \infty$*, a coupling* $(\hat{\mathbf{Z}}_1^{k_1}, \hat{\mathbf{Z}}_2^{k_2}, \ldots, \hat{\mathbf{Z}})$ *of* $\mathbf{Z}_1^{k_1}, \mathbf{Z}_2^{k_2}, \ldots, \mathbf{Z}$ *and an* $\mathbb{N}$*-valued* N *such that*

$$\hat{\mathbf{Z}}_n^{k_n} = \hat{\mathbf{Z}}^{k_n}, \qquad n \geq N.$$

▷

Proof. First, assume the coupling claim. Fix $k \in \mathbb{N}$, take $m \in \mathbb{N}$ such that $k_m \geq k$, and note that then the coupling claim yields $\hat{\mathbf{Z}}_n^k = \hat{\mathbf{Z}}^k$ for $n \geq \max\{N, m\}$. Theorem 2 now yields the density claim.

Conversely, with Q, Q^k, Q_n^k the distributions of $\mathbf{Z}$, $\mathbf{Z}^k$, $\mathbf{Z}_n^k$ and with f_n^k a density of Q_n^k, assume that $g_n^k := \inf_{i \geq n} f_i^k \nearrow$ to a density of Q^k, $n \to \infty$. Hence with ν_n^k the measure with density g_n^k,

$$\nu_1^k \leq \nu_2^k \leq \cdots \nearrow Q^k, \quad k \in \mathbb{N}.$$

Thus there are numbers $0 \leq n_1 \leq n_2 \leq \cdots \to \infty$ such that

$$0 \leq Q^k - \nu_{n_k}^k \leq 2^{-k}, \quad k \in \mathbb{N}_0.$$

For $A \in \mathcal{H}$ and $\mathbf{z}^k \in H^k$, let $q_k(A \,|\, \mathbf{z}^k)$ be the conditional probability of the event $\{\mathbf{Z} \in A\}$ given $\mathbf{Z}^k = \mathbf{z}^k$. Then

$$Q(A) = \textstyle\int q_k(A \,|\, \cdot)\, dQ^k, \quad A \in \mathcal{H}.$$

Use $\nu_{n_k}^k \ll Q^k$ to extend $\nu_{n_k}^k$ from $(H^k, \mathcal{H}^k)$ to a ν_k on $(H, \mathcal{H})$ by

$$\nu_k(A) := \textstyle\int q_k(A \,|\, \cdot)\, d\nu_{n_k}^k, \quad A \in \mathcal{H}.$$

For $k \in \mathbb{N}_0$ the last three displays yield

$$0 \leq Q - \nu_k \leq 2^{-k}. \tag{1}$$

Let h_k be a density of ν_k with respect to Q. For integers $k < m$ let $\nu_{k,m}$ be the measure with density $\min_{k \leq j \leq m} h_j$ with respect to Q. Partition H into sets $A_k, ..., A_m \in \mathcal{H}$ such that $\min_{k \leq j \leq m} h_j = h_i$ on A_i and thus

$$\nu_{k,m}(\cdot \cap A_i) = \nu_i(\cdot \cap A_i), \quad k \leq i \leq m.$$

This and (1) yield

$$0 \leq Q - \nu_{k,m} = \sum_{k \leq i \leq m} \big(Q(\cdot \cap A_i) - \nu_i(\cdot \cap A_i)\big) \leq \sum_{k \leq i < \infty} 2^{-i} = 2^{-k+1}.$$

Define $k_n := k$ if $n_k \leq n < n_{k+1}$. Then $0 \leq Q - \nu_{k_n,m} \leq 2^{-k_n+1}$. Let μ_n be the measure with density $\inf_{i \geq k_n} h_i$ with respect to Q and send $m \to \infty$ to obtain $0 \leq Q - \mu_n \leq 2^{-k_n+1}$. Thus

$$0 =: \mu_0 \leq \mu_1 \leq \mu_2 \leq \cdots \nearrow Q.$$

Since $g_n^k \leq f_n^k$ and $\inf_{i \geq k_n} h_i \leq h_{k_n}$ we have $\nu_n^k \leq Q_n^k$ and $\mu_n \leq \nu_{k_n}$. Thus

$$\mu_n^{k_n} \leq Q_n^{k_n}, \quad n \in \mathbb{N}.$$

Now apply the last two displays: let $(\Omega, \mathcal{F}, \mathbb{P})$ support independent random elements N, $\mathbf{V}_1$, $\mathbf{V}_2, \ldots, \mathbf{W}_1$, $\mathbf{W}_2, \ldots$ such that

N is $\mathbb{N}$ valued and $\mathbb{P}(N \leq n) = \mu_n(H)$, $n \in \mathbb{N}$,

$\mathbf{V}_n$ is H valued and $\mathbb{P}(\mathbf{V}_n \in \cdot\,) = (\mu_n - \mu_{n-1})/\mathbb{P}(N = n)$,

$\mathbf{W}_n$ is H^{k_n} valued and $\mathbb{P}(\mathbf{W}_n \in \cdot\,) = (Q_n^{k_n} - \mu_n^{k_n})/\mathbb{P}(N > n)$.

Put $\hat{\mathbf{Z}}_n^{k_n} = \mathbf{V}_N^{k_n}$ on $\{N \le n\}$ and $\hat{\mathbf{Z}}_n^{k_n} = \mathbf{W}_n$ on $\{N > n\}$. Then

$$\begin{aligned}\mathbb{P}(\hat{\mathbf{Z}}_n^{k_n} \in \cdot) &= \sum_{i=1}^{n} \mathbb{P}(\mathbf{V}_i^{k_n} \in \cdot)\mathbb{P}(N=i) + \mathbb{P}(\mathbf{W}_n \in \cdot)\mathbb{P}(N>n) \\ &= \sum_{i=1}^{n} (\mu_i^{k_n} - \mu_{i-1}^{k_n}) + (Q_n^{k_n} - \mu_n^{k_n}) = Q_n^{k_n}.\end{aligned}$$

Put $\hat{\mathbf{Z}} = \mathbf{V}_N$. Then

$$\mathbb{P}(\hat{\mathbf{Z}} \in \cdot) = \sum_{i=1}^{\infty} \mathbb{P}(\mathbf{V}_i \in \cdot)\mathbb{P}(N=i) = \sum_{i=1}^{\infty} (\mu_i - \mu_{i-1}) = Q.$$

Clearly $\hat{\mathbf{Z}}_n^{k_n} = \hat{\mathbf{Z}}^{k_n}, \quad n \ge N.$ □

We now extend $(\hat{\mathbf{Z}}_1^{k_1}, \hat{\mathbf{Z}}_2^{k_2}, \dots, \hat{\mathbf{Z}})$ to a full coupling $(\hat{\mathbf{Z}}_1, \hat{\mathbf{Z}}_2, \dots, \hat{\mathbf{Z}})$ in a special case. This extension will be used in the next section.

Corollary. *Let* $(H, \mathcal{H}) := (\mathbb{N}, 2^{\mathbb{N}}) \otimes (\mathbb{N}, 2^{\mathbb{N}}) \otimes \cdots \otimes (E, \mathcal{E})$. *Then*

$$\forall k \in \mathbb{N} \text{ and } \mathbf{i}^k \in \mathbb{N}^k: \quad \mathbb{P}_n(\mathbf{Z}_n^k = \mathbf{i}^k) \to \mathbb{P}(\mathbf{Z}^k = \mathbf{i}^k) \quad \text{as} \quad n \to \infty$$

if and only if there exists a coupling $(\hat{\mathbf{Z}}_1, \hat{\mathbf{Z}}_2, \dots, \hat{\mathbf{Z}})$ *of* $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}$*, an* $\mathbb{N}$*-valued random variable* N *and integers* $0 \le k_1 \le k_2 \le \cdots \to \infty$ *such that*

$$\hat{\mathbf{Z}}_n^{k_n} = \hat{\mathbf{Z}}^{k_n}, \qquad n \ge N$$

Proof. Let $\mathbf{V}_{n,\mathbf{i}^{k_n}}$, $n \in \mathbb{N}$, $\mathbf{i}^{k_n} \in \mathbb{N}^{k_n}$, be random elements that are independent of $(\hat{\mathbf{Z}}_1^{k_1}, \hat{\mathbf{Z}}_2^{k_2}, \dots, \hat{\mathbf{Z}}, N)$ and such that $\mathbf{V}_{n,\mathbf{i}^{k_n}}^{k_n} = \mathbf{i}^{k_n}$ and $\mathbb{P}(\mathbf{V}_{n,\mathbf{i}^{k_n}} \in \cdot) = \mathbb{P}(\mathbf{Z}_n \in \cdot \,|\, \mathbf{Z}_n^{k_n} = \mathbf{i}^{k_n})$. Since $\mathbf{V}_{n,\hat{\mathbf{Z}}_n^{k_n}}^{k_n} = \hat{\mathbf{Z}}_n^{k_n}$ we can extend $\hat{\mathbf{Z}}_n^{k_n}$ consistently to a $\hat{\mathbf{Z}}_n$ by $\hat{\mathbf{Z}}_n := \mathbf{V}_{n,\hat{\mathbf{Z}}_n^{k_n}}$. □

Remark. If $(E, \mathcal{E})$ is Polish then $(\mathbb{N}, 2^{\mathbb{N}})$ can be replaced by a Polish space. Moreover, Theorem 3 holds for general continuous-time $\hat{\mathbf{Z}}$, and if $(E, \mathcal{E})$ is Polish then in the corollary the discrete-time processes can be replaced by processes in Skorohod space. For proofs of these results, see [4].

4. Skorohod Representation Theorem – Proof

The if-direction of Theorem 1 is straightforward: if h is continuous and bounded and $\hat{X}_n \to \hat{X}$ pointwise as $n \to \infty$ then also $h(\hat{X}_n) \to h(\hat{X})$, which by bounded convergence yields $\int h dP_n \to \int h dP$ as $n \to \infty$.

In order to prove the only-if-direction of Theorem 1, assume from now on that $X_n \to X$ in distribution as $n \to \infty$. A set $A \in \mathcal{E}$ is called a P-continuity set if its boundary has P-mass 0. According to the Portmanteau Theorem, convergence in distribution is equivalent to

$$P_n(A) \to P(A), \; n \to \infty, \text{ for all } P\text{-continuity sets } A.$$

Assume also that there is a separable set $S \in \mathcal{E}$ such that $P(S) = 1$.

Partition E recursively into nested partitions of P-continuity sets:

Let $A_2, A_3, \dots$ be disjoint P-continuity sets covering S, of diameters < 1. Put $A_1 = E \setminus (A_2 \cup A_3 \cup \dots)$ and note that A_1 is also a P-continuity set.

Put $A_{11} = A_1$ and $A_{12} = A_{13} = \dots = \emptyset$. For $i > 1$, let $A_{i2}, A_{i3}, \dots$ be disjoint P-continuity sets covering $S \cap A_i$, each of diameter $< 1/2$. Put $A_{i1} = A_i \setminus (A_{i2} \cup A_{i3} \cup \dots)$, note that A_{i1} is also a P-continuity set.

Continue this recursively to obtain that

$$\{A_{\mathbf{i}^k} : \mathbf{i}^k \in \mathbb{N}^k\},\ k \in \mathbb{N}, \text{ are nested } E\text{-partitions of } P\text{-continuity sets,} \\ \text{and } A_{\mathbf{i}^k},\ \mathbf{i}^k \in (\mathbb{N} \setminus \{1\})^k, \text{ cover } S \text{ and are of diameters } < 1/k. \quad (2)$$

Put

$$\mathbf{Z}_n = (Z_n^1, Z_n^2, \dots, X_n) \quad \text{and} \quad \mathbf{Z} = (Z^1, Z^2, \dots, X)$$

where Z_n^k and Z^k are defined as follows: for $k \in \mathbb{N}$, $\mathbf{i}^k \in \mathbb{N}^k$,

$$\mathbf{Z}_n^k = \mathbf{i}^k \text{ if } X_n \in A_{\mathbf{i}^k} \qquad \text{and} \qquad \mathbf{Z}^k = \mathbf{i}^k \text{ if } X \in A_{\mathbf{i}^k}.$$

Now $P_n(A_{\mathbf{i}^k}) \to P(A_{\mathbf{i}^k})$ yields $\mathbb{P}(\mathbf{Z}_n^k = \mathbf{i}^k) \to \mathbb{P}(\mathbf{Z}^k = \mathbf{i}^k)$ as $n \to \infty$. Apply the corollary to obtain a coupling $(\hat{\mathbf{Z}}_1, \hat{\mathbf{Z}}_2, \dots, \hat{\mathbf{Z}})$ and an $\mathbb{N}$-valued N such that $\hat{\mathbf{Z}}_n^{k_n} = \hat{\mathbf{Z}}^{k_n}$ for $n \geq N$. Delete a null-event to obtain from this that $\hat{X} \in S$ and that for $k \in \mathbb{N}$, $\mathbf{i}^k \in \mathbb{N}^k$,

$$\hat{\mathbf{Z}}_n^k = \mathbf{i}^k \text{ if } \hat{X}_n \in A_{\mathbf{i}^k} \qquad \text{and} \qquad \hat{\mathbf{Z}}^k = \mathbf{i}^k \text{ if } \hat{X} \in A_{\mathbf{i}^k}.$$

Thus $\hat{X}_n \in A_{\hat{\mathbf{Z}}_n^{k_n}}$ and $\hat{X} \in A_{\hat{\mathbf{Z}}^{k_n}}$ for all n. Now $\hat{\mathbf{Z}}_n^{k_n} = \hat{\mathbf{Z}}^{k_n}$ for $n \geq N$ so

$$\text{for } n \geq N: \qquad \text{both } \hat{X}_n \text{ and } \hat{X} \in A_{\hat{\mathbf{Z}}^{k_n}}.$$

Since $\hat{X} \in S$ we have $\hat{\mathbf{Z}}^{k_n} \in (\mathbb{N} \setminus \{1\})^{k_n}$ and (2) yields (with d the metric)

$$\text{for } n \geq N: \qquad d(\hat{X}_n, \hat{X}) < 1/k_n \to 0 \text{ as } n \to \infty. \qquad \square$$

References

1. *Dudley R.M.* Distances of probability measures and random variables // Ann. Math. Statist. — 1958. — Vol. 39. — P. 1563–1572.
2. *Skorohod A.V.* Limit theorems for stochastic processes // Theo. Probab. Appl. — 1956. — Vol. 1. — P. 261–290.
3. *Thorisson H.* Coupling methods in probability theory // Scand. J. Statist. — 1995. — Vol. 22. — P. 159–182.
4. *Thorisson H.* Convergence in Density in Finite Time Windows and the Skorohod Representation // Electron. Commun. Probab. — 2016. — Vol. 21, no. 63. — P. 1–9.
5. *Wichura M.J.* On the construction of almost uniformly convergent random variables with given weakly convergent image laws // Ann. Math. Statist. — 1970. — Vol. 41. — P. 284–291.