3 Coupling

Two random variables, say X and Y, are coupled, if they are defined on the same probablity space. To couple two given variables X and Y, one usually defines a random vector $\big(\widetilde{X}, \widetilde{Y}\big)$ with joint probability $\widetilde{\mathsf{P}}(\cdot,\cdot)$ on some probability space[18] so that the marginal distribution of $\widetilde{X}$ coincides with the distribution of X and the marginal distribution of $\widetilde{Y}$ coincides with the distribution of Y.

Example 3.1. *Fix p_1, $p_2 \in [0,1]$ such that $p_1 \le p_2$ and consider the following joint distributions (we write $q_i = 1 - p_i$):*

	0	1	$\widetilde{X}$
0	$q_1 q_2$	$q_1 p_2$	q_1
1	$p_1 q_2$	$p_1 p_2$	p_1
$\widetilde{Y}$	q_2	p_2	

	0	1	$\widetilde{X}$
0	q_2	$p_2 - p_1$	q_1
1	0	p_1	p_1
$\widetilde{Y}$	q_2	p_2	

It is easy to see that in both cases[19] $\widetilde{X} \sim \mathsf{Ber}(p_1)$, $\widetilde{Y} \sim \mathsf{Ber}(p_2)$, *though in the first case $\widetilde{X}$ and $\widetilde{Y}$ are independent, whereas in the second case we have* $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$.

3.1 Stochastic domination

If $X \sim \mathsf{Ber}(p)$, its tail probabilities $\mathsf{P}(X > a)$ satisfy

$$\mathsf{P}(X > a) = \begin{cases} 1, & a < 0\,, \\ p, & 0 \le a < 1\,, \\ 0, & a \ge 1\,. \end{cases}$$

Consequently, in the setup of Example 3.1, for the variable $X \sim \mathsf{Ber}(p_1)$ and $Y \sim \mathsf{Ber}(p_2)$ with $p_1 \le p_2$ we have $\mathsf{P}(X > a) \le \mathsf{P}(Y > a)$ for all $a \in \mathbb{R}$. The last inequality is useful enough to deserve a name:

✍ **Definition 3.2.** [Stochastic domination] A random variable X is *stochastically smaller* than a random variable Y (write $X \preccurlyeq Y$) if the inequality

$$\mathsf{P}\big(X > x\big) \le \mathsf{P}\big(Y > x\big) \tag{3.1}$$

holds for all $x \in \mathbb{R}$.

☞ **Remark 3.2.1.** If $X \preccurlyeq Y$ and $g(\,\cdot\,) \ge 0$ is an arbitrary increasing function on $\mathbb{R}$, then $g(X) \preccurlyeq g(Y)$. If, in addition, $X \ge 0$, $Y \ge 0$, and $g(\cdot)$ is smooth with $g(0) = 0$, then

$$\mathsf{E}g(X) \equiv \int_0^\infty g'(z)\mathsf{P}(X > z)\,dz \le \int_0^\infty g'(z)\mathsf{P}(Y > z)\,dz \equiv \mathsf{E}g(Y)\,.$$

☞ [18] A priori the original variables X and Y can be defined on arbitrary probability spaces, so that one has no reason to expect that these spaces can be "joined" in any way!

[19] and in fact, every convex linear combination of these two tables provides a joint distribution with the same marginals.

Exercise 3.3. Generalise the inequality above to a broader class of functions and verify that if $X \preccurlyeq Y$, then $\mathsf{E}(X^{2k+1}) \le \mathsf{E}(Y^{2k+1})$, $\mathsf{E}e^X \le \mathsf{E}e^Y$ etc.

In the setup of Example 3.1, if $X \sim \mathsf{Ber}(p_1)$ and $Y \sim \mathsf{Ber}(p_2)$ then X is stochastically smaller than Y (ie., $X \preccurlyeq Y$) if and only if $p_1 \le p_2$; moreover, this is equivalent to existence of a coupling $(\widetilde{X}, \widetilde{Y})$ of X and Y in which these variables are ordered with probablity one, $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$. The next result shows that this is a rather generic situation:

✍ **Lemma 3.4.** *A random variable X is stochastically smaller than a random variable Y if and only if there exists a coupling $(\widetilde{X}, \widetilde{Y})$ of X and Y such that $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$.*

Remark 3.4.1. Notice that one claim of Lemma 3.4 is immediate from

$$\mathsf{P}(x < X) \equiv \widetilde{\mathsf{P}}\big(x < \widetilde{X}\big) = \widetilde{\mathsf{P}}\big(x < \widetilde{X} \le \widetilde{Y}\big) \le \widetilde{\mathsf{P}}\big(x < \widetilde{Y}\big) \equiv \mathsf{P}\big(x < Y\big)\,;$$

the other claim requires a more advanced argument (we shall not do it here!).

Example 3.5. *If $X \sim \mathsf{Bin}(m,p)$ and $Y \sim \mathsf{Bin}(n,p)$ with $m \le n$, then $X \preccurlyeq Y$.*

Solution. Let $Z_1 \sim \mathsf{Bin}(m,p)$ and $Z_2 \sim \mathsf{Bin}(n-m,p)$ be independent variables defined on the same probability space. We then put $\widetilde{X} = Z_1$ and $\widetilde{Y} = Z_1 + Z_2$ so that $\widetilde{Y} - \widetilde{X} = Z_2 \ge 0$ with probability one, $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$, and $X \sim \widetilde{X}$, $Y \sim \widetilde{Y}$. □

Example 3.6. *If $X \sim \mathsf{Poi}(\lambda)$ and $Y \sim \mathsf{Poi}(\mu)$ with $\lambda \le \mu$, then $X \preccurlyeq Y$.*

Solution. Let $Z_1 \sim \mathsf{Poi}(\lambda)$ and $Z_2 \sim \mathsf{Poi}(\mu - \lambda)$ be independent variables defined on the same probability space.[20] We then put $\widetilde{X} = Z_1$ and $\widetilde{Y} = Z_1 + Z_2$ so that $\widetilde{Y} - \widetilde{X} = Z_2 \ge 0$ with probability one, $\widetilde{\mathsf{P}}(\widetilde{X} \le \widetilde{Y}) = 1$, and $X \sim \widetilde{X}$, $Y \sim \widetilde{Y}$. □

☞ **Example 3.7.** *Let $(X_n)_{n\ge0}$ be a branching process with offspring distribution $\{p_m\}_{m\ge0}$ and $X_0 = 1$. Let $(Y_n)_{n\ge0}$ be a branching process with the same offspring distribution and $Y_0 = 2$. Show that $\mathsf{P}(X_n = 0) \ge \mathsf{P}(Y_n = 0)$.*

Solution. It is enough to show that $X_n \preccurlyeq Y_n$ for all $n \ge 0$. To this end consider two independent branching processes, $(Z'_n)_{n\ge0}$ and $(Z''_n)_{n\ge0}$, having the same offspring distribution $\{p_m\}_{m\ge0}$ and satisfying $Z'_0 = Z''_0 = 1$. We then put $\widetilde{X}_n = Z'_n$ and $\widetilde{Y}_n = Z'_n + Z''_n$ for all $n \ge 0$, so that $\widetilde{Y}_n - \widetilde{X}_n = Z''_n \ge 0$, ie., $\widetilde{\mathsf{P}}(\widetilde{X}_n \le \widetilde{Y}_n) = 1$ and $X_n \sim \widetilde{X}_n$, $Y_n \sim \widetilde{Y}_n$ for all $n \ge 0$. □

Exercise 3.8. For an offspring distribution $\{p_m\}_{m\ge0}$, let $(X_n)_{n\ge0}$ be the branching process with $X_0 = k$ and let $(Y_n)_{n\ge0}$ be the branching process with $Y_0 = l$, $k < l$. Show that $X_n \preccurlyeq Y_n$ for all $n \ge 0$.

Exercise 3.9. Let $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$ be standard branching processes with $X_0 = Y_0 = 1$. Assume that the offspring distribution of X is stochastically smaller than that of Y, ie., $X_1 \preccurlyeq Y_1$. Show that $X_n \preccurlyeq Y_n$ for all $n \ge 0$.

[20] here and below we assume that $Z \sim \mathsf{Poi}(0)$ means that $\mathsf{P}(Z = 0) = 1$.

Exercise 3.10. Let $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ be two branching processes, such that the offspring distribution of X is stochastically smaller than that of Y, ie., for all integer $k \geq 0$, $\mathsf{P}(X_1 > k | X_0 = 1) \leq \mathsf{P}(Y_1 > k | Y_0 = 1)$. If for positive integers m, $l > m$, we have $X_0 = m < l = Y_0$, show that[21] $X_n \preccurlyeq Y_n$ for all $n \geq 0$.

Exercise 3.11. Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ be Gaussian r.v.'s.
a) If $\mu_X \leq \mu_Y$ but $\sigma_X^2 = \sigma_Y^2$, is it true that $X \preccurlyeq Y$?
b) If $\mu_X = \mu_Y$ but $\sigma_X^2 \leq \sigma_Y^2$, is it true that $X \preccurlyeq Y$?

Exercise 3.12. Let X and Y be two exponential random variables, $X \sim \mathsf{Exp}(\lambda)$ and $Y \sim \mathsf{Exp}(\mu)$. If $0 < \lambda \leq \mu < \infty$, are the variables X and Y stochastically ordered? Justify your answer.[22]

Exercise 3.13. Let X and Y be two geometric random variables, $X \sim \mathsf{Geom}(p)$ and $Y \sim \mathsf{Geom}(r)$. If $0 < p \leq r < 1$, are the variables X and Y stochastically ordered? Justify your answer.

Exercise 3.14. Let X and Y be two gamma random variables,[23] $X \sim \Gamma(a, \lambda)$ and $Y \sim \Gamma(b, \lambda)$. If $0 < a \leq b < \infty$, are the variables X and Y stochastically ordered? Justify your answer.
[**Hint:** If $Z_1 \sim \Gamma(c_1, \lambda)$ and $Z_2 \sim \Gamma(c_2, \lambda)$ are independent, then $Z_1 + Z_2 \sim \Gamma(c_1 + c_2, \lambda)$.]

3.2 Total variation distance

✍ **Definition 3.15.** [Total Variation Distance] Let μ and ν be two probability measures on the same probability space. The *total variation* distance between μ and ν is

$$\mathsf{d_{TV}}(\mu, \nu) \stackrel{\mathsf{def}}{=} \max_A \bigl|\mu(A) - \nu(A)\bigr| . \tag{3.2}$$

☞ **Exercise 3.16.** Let μ have p.m.f. $\{p_x\}$ and let ν have p.m.f. $\{q_y\}$. Show that then

$$\mathsf{d_{TV}}(\mu, \nu) \equiv \mathsf{d_{TV}}\bigl(\{p\}, \{q\}\bigr) = \frac{1}{2} \sum_z \bigl|p_z - q_z\bigr| .$$

Deduce that $\mathsf{d_{TV}}(\cdot, \cdot)$ is a distance between probability measures[24] (ie., it is non-negative, symmetric, and satisfies the triangle inequality) such that $\mathsf{d_{TV}}(\cdot, \cdot) \leq 1$ for all probability measures μ and ν.
[**Hint:** The equality (3.2) is saturated for $A = \{x : p_x > q_x\}$.]

An important relation between coupling and the total variation distance is explained by the following fact.

[21] this result can also be generalised to the case where X_0 and Y_0 are stochastically ordered *random* variables, ie., $X_0 \preccurlyeq Y_0$.
[22] by proving the result or giving a counter-example!
[23] The density of $Z \sim \Gamma(a, \lambda)$ is $\frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$ $(x > 0)$; notice that $\Gamma(1, \lambda)$ is just $\mathsf{Exp}(\lambda)$.
[24] so that all probability measures form a metric space for this distance!

✍ **Example 3.17.** [Maximal Coupling] *Let random variables X and Y be such that* $\mathsf{P}(X=x)=p_x$ *and* $\mathsf{P}(Y=y)=q_y$. *Define the coupling* $\big(\widetilde{X},\widetilde{Y}\big)$ *via*[25]

$$\begin{gathered}\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=z\big)=\min\big(p_z,q_z\big)\,,\\ \widehat{\mathsf{P}}\big(\widetilde{X}=x,\widetilde{Y}=y\big)=\frac{\big(p_x-\min(p_x,q_x)\big)\,\big(q_y-\min(p_y,q_y)\big)}{\mathsf{d_{TV}}\big(\{p\},\{q\}\big)}\,,\quad x\neq y\,.\end{gathered}\tag{3.3}$$

Exercise 3.18. Show that

$$\sum_x\big(p_x-\min(p_x,q_x)\big)=\sum_y\big(q_y-\min(p_y,q_y)\big)=\mathsf{d_{TV}}\big(\{p\},\{q\}\big)\,,$$

and deduce that (3.3) is indeed a coupling of X and Y (ie., that (3.3) defines a probability distribution with correct marginals).

☞ **Example 3.19.** *Consider* $X\sim\mathsf{Ber}(p_1)$ *and* $Y\sim\mathsf{Ber}(p_2)$ *with* $p_1\le p_2$. *It is a straightforward exercise to check that the second table in Example 3.1 provides the maximal coupling of X and Y. We notice also that in this case*

$$\widehat{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)=p_2-p_1=\mathsf{d_{TV}}(X,Y)\,.$$

Lemma 3.20. *Let* $\widehat{\mathsf{P}}(\cdot,\cdot)$ *be the maximal coupling of X and Y as defined in (3.3). Then for every other coupling* $\widetilde{\mathsf{P}}(\cdot\,,\,\cdot)$ *of X and Y we have*

$$\widetilde{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)\ge\widehat{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)=\mathsf{d_{TV}}\big(\widetilde{X},\widetilde{Y}\big)\,.\tag{3.4}$$

Proof. Summing the inequalities $\widetilde{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=z\big)\le\min(p_z,q_z)$ we deduce

$$\widetilde{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)\ge1-\sum_z\min(p_z,q_z)=\sum_z\big(p_z-\min(p_z,q_z)\big)=\mathsf{d_{TV}}\big(\{p\},\{q\}\big)\,,$$

in view of Exercise 3.18 and Example 3.19. □

Remark 3.20.1. Notice that according to (3.4),

$$\widetilde{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}\big)\le\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}\big)=1-\mathsf{d_{TV}}\big(\widetilde{X},\widetilde{Y}\big)\,,$$

ie., the probability that $\widetilde{X}=\widetilde{Y}$ is maximised under the optimal coupling $\widehat{\mathsf{P}}(\cdot\,,\,\cdot)$.

☞ **Example 3.21.** *The maximal coupling of* $X\sim\mathsf{Ber}(p)$ *and* $Y\sim\mathsf{Poi}(p)$ *satisfies*

$$\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}=x\big)=\begin{cases}1-p\,, & x=0\,,\\ pe^{-p}\,, & x=1\,,\\ 0\,, & x>1\,,\end{cases}$$

so that

$$\mathsf{d_{TV}}\big(\widetilde{X},\widetilde{Y}\big)\equiv\widehat{\mathsf{P}}\big(\widetilde{X}\neq\widetilde{Y}\big)=1-\widehat{\mathsf{P}}\big(\widetilde{X}=\widetilde{Y}\big)=p\big(1-e^{-p}\big)\le p^2\,.\tag{3.5}$$

Is any of the variables X and Y stochastically dominated by another?

[25] By Exercise 3.16, if $\mathsf{d_{TV}}\big(\{p\},\{q\}\big)=0$, we have $p_z=q_z$ for all z, and thus all off-diagonal terms in (3.3) vanish.

3.3 Applications to convergence

3.3.1 The Law of rare events

The following result provides an alternative derivation of the convergence result in Exercise 1.22.

Theorem 3.22. *Let $X = \sum_{k=1}^n X_k$, where $X_k \sim \mathsf{Ber}(p_k)$ are independent random variables. Let, further, $Y \sim \mathsf{Poi}(\lambda)$, where $\lambda = \sum_{k=1}^n p_k$. Then the maximal coupling of X and Y satisfies*

$$\mathsf{d}_{\mathsf{TV}}\big(\widetilde{X},\widetilde{Y}\big) \equiv \widehat{\mathsf{P}}\big(\widetilde{X} \neq \widetilde{Y}\big) \le \sum_{k=1}^n \big(p_k\big)^2 .$$

Proof. Write $Y = \sum_{k=1}^n Y_k$, where $Y_k \sim \mathsf{Poi}(p_k)$ are independent rv's. Of course, $\mathsf{P}\big(\sum_{k=1}^n X_k \neq \sum_{k=1}^n Y_k\big) \le \sum_{k=1}^n \mathsf{P}(X_k \neq Y_k)$ for every joint distribution of $(X_k)_{k=1}^n$ and $(Y_k)_{k=1}^n$. Let $\widehat{\mathsf{P}}_k$ be the maximal coupling for the pair $\{X_k, Y_k\}$, and let $\widehat{\mathsf{P}}_0$ be the maximal coupling for two sums. Notice that the LHS above is not smaller than $\mathsf{d}_{\mathsf{TV}}\big(\widetilde{X},\widetilde{Y}\big) \equiv \widehat{\mathsf{P}}_0\big(\widetilde{X} \neq \widetilde{Y}\big)$; on the other hand, using the (independent) product measure $\mathsf{P} = \widehat{\mathsf{P}}_1 \times \cdots \times \widehat{\mathsf{P}}_n$ on the right we deduce that then the RHS becomes just $\sum_{k=1}^n \widehat{\mathsf{P}}_k\big(\widetilde{X}_k \neq \widetilde{Y}_k\big)$. The result now follows from (3.5). □

Exercise 3.23. Let $X \sim \mathsf{Bin}(n, \frac{\lambda}{n})$ and $Y \sim \mathsf{Poi}(\lambda)$ for some $\lambda > 0$. Show that

$$\frac{1}{2}\big|\mathsf{P}(X = k) - \mathsf{P}(Y = k)\big| \le \mathsf{d}_{\mathsf{TV}}\big(\widetilde{X},\widetilde{Y}\big) \le \frac{\lambda^2}{n} \qquad \text{for every } k \ge 0. \qquad (3.6)$$

Deduce that for every fixed $k \ge 0$, we have $\mathsf{P}(X = k) \to \frac{\lambda^k}{k!}e^{-\lambda}$ as $n \to \infty$.

☞ **Remark 3.23.1.** Notice that (3.6) implies that if $X \sim \mathsf{Bin}(n,p)$ and $Y \sim \mathsf{Poi}(np)$ then for every $k \ge 0$ the probabilities $\mathsf{P}(X = k)$ and $\mathsf{P}(Y = k)$ differ by at most $2np^2$. In particular, if $n = 10$ and $p = 0.01$ the discrepancy between any pair of such probabilities is bounded above by 0.002, ie., they coincide in the first two decimal places.

3.3.2 Coupling of Markov chains

Coupling is a very useful tool for proving convergence towards equilibrium for various processes, including Markov chains and random walks. The following example uses the independent coupling[26] of two random walks on a complete graph.

☞ **Example 3.24.** *On the complete graph K_m on m vertices, consider the "lazy" random walk $(X_n)_{n\ge0}$ with $X_0 = x$, such that at every step it jumps to any of the vertices of K_m (including the current one) uniformly at random. To show that eventually $(X_n)_{n\ge0}$ forgets its initial distribution, one couples $(X_n)_{n\ge0}$ with another random walk $(Y_n)_{n\ge0}$, $Y_0 = y$, of the same type, and shows that for large n both processes coincide with large probability.*

[26] Recall the first table in Example 3.1.

Solution. We treat the pair $(\widetilde{X}_n, \widetilde{Y}_n)_{n\ge0}$ as a two-component random walk on $K_m \times K_m$ with the following jump probabilities: the transition $(x,y) \mapsto (x',y')$ occurs with probability

$$\tilde{p}_{(x,y)(x',y')} = \begin{cases} \frac{1}{m^2}\,, & \text{if } x \ne y\,,\\ \frac{1}{m}\,, & \text{if } x = y \text{ and } x' = y'\,,\\ 0\,, & \text{if } x = y \text{ and } x' \ne y'\,. \end{cases}$$

It is straightforward to check that this is a correct coupling:[27]

$$\sum_{y'} \tilde{p}_{(x,y)(x',y')} = m \cdot \frac{1}{m^2} = p_{x,x'}\,, \qquad \sum_{y'} \tilde{p}_{(x,x)(x',y')} = \tilde{p}_{(x,x)(x',x')} = \frac{1}{m} = p_{x,x'}\,.$$

☞ Notice that the walks $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$ run independently until they meet, and from that moment onwards they move together. For $x \ne y$, denote

$$T \stackrel{\mathsf{def}}{=} \min\{n \ge 1 : X_n = Y_n\}\,. \tag{3.7}$$

By a straightforward induction, it is easy to see that $\mathsf{P}(T = k) = \frac{1}{m}\big(1 - \frac{1}{m}\big)^{k-1}$, ie., T is a geometric random variable. Since X_n and Y_n coincide after they meet, we have $\mathsf{P}(X_n = j, T \le n) = \mathsf{P}(Y_n = j, T \le n)$ for all vertices j in K_m; therefore

$$\mathsf{P}\big(X_n = j\big) - \mathsf{P}\big(Y_n = j\big) = \mathsf{P}(X_n = j, T > n) - \mathsf{P}(Y_n = j, T > n)\,.$$

As a result,

$$\begin{aligned} \mathsf{d}_{\mathsf{TV}}(X_n, Y_n) &= \frac{1}{2}\sum_j \big|\mathsf{P}\big(X_n = j\big) - \mathsf{P}\big(Y_n = j\big)\big| \\ &\le \frac{1}{2}\sum_j \Big(\mathsf{P}\big(X_n = j, T > n\big) + \mathsf{P}\big(Y_n = j, T > n\big)\Big) = \mathsf{P}(T > n)\,, \end{aligned}$$

and it remains to observe that the RHS above is exponentially small, $\mathsf{P}(T > n) \le \big(1 - \frac{1}{m}\big)^n \le e^{-n/m}$; in other words, the random walk $(X_n)_{n\ge0}$ forgets its initial state $X_0 = x$ at least exponentially fast.

Notice that if $(Y_n)_{n\ge0}$ starts from the equilibrium (ie., its initial position is selected in K_m uniformly at random), our argument shows that for every initial state x and every state j we have $|\mathsf{P}(X_n = j) - \frac{1}{m}| \le e^{-n/m}$, ie., convergence towards the equilibrium distribution is at least exponentially fast.[28] □

☞ **Exercise 3.25.** Let $(X_n)_{n\ge0}$ be a random walk on the complete graph K_m (ie., the Markov chain jumping from each vertex to each of its neighbours with equal probability). Carefully define the maximal coupling for this random walk and use it to get the best bound you can on the speed of convergence towards the equilibrium distribution.

Now do the same for the "lazy" version of the walk, recall Example 3.24. Explain what happens.

[27] the argument for the other marginal is similar;

[28] Exercise 3.25 gives a more precise information about this convergence.

Example 3.26. [Random walk on a hypercube[29]] *An n-dimensional hypercube is just $\mathcal{H}_n \equiv \{0,1\}^n$, a graph whose vertex set is the collection of all n-sequences made of 0 and 1 (there are 2^n of them) and whose edges connect vertices which differ at a single position. The "lazy" random walk $(W_k)_{k\ge0}$ on $\mathcal{H}_n$ is defined as follows: if $W_k = v \in \mathcal{H}_n$, select m, $1 \le m \le n$, uniformly at random and flip a fair coin. If the coin shows heads, set the mth coordinate of v to 1, otherwise set it to 0. The walk $(W_k)_{k\ge0}$ is aperiodic and irreducible; as k increases, it tends to forget its initial state.*

Exercise 3.27.⊛ Use the argument of Example 3.24 and describe the convergence towards the equilibrium for the random walk $(W_k)_{k\ge0}$ on the hypercube $\mathcal{H}_n$.
[**Hint:** Show that the coupling time T defined as in (3.7), has the same distribution as the time until success in the coupon collector's problem.]

☞ **Exercise 3.28.**⊛ Let $(\xi_j)_{j\ge1}$ be i.i.d.r.v.'s such that $\mathsf{P}(\xi=1)=1-\mathsf{P}(\xi=-1)=p_x$ and let $(\eta_j)_{j\ge1}$ be i.i.d.r.v.'s such that $\mathsf{P}(\eta=1)=1-\mathsf{P}(\eta=-1)=p_y$. Define simple random walks $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$ on $\mathbb{Z}$ via $X_n = x+\sum_{j=1}^n \xi_j$ and $Y_n = y+\sum_{j=1}^n \eta_j$.

a) Show that the random walk $(X_n)_{n\ge0}$ on $\mathbb{Z}$ forgets its initial state x; namely, for $p_x = p_y$ and $y-x=2k$, construct a coupling of $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$ similar to that of Example 3.24.

b) Show that the random walk $(X_n)_{n\ge0}$ monotonously depends on p_x; namely, for $x\le y$ and $p_x\le p_y$, use the ideas from Example 3.1 to construct a monotone coupling of $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$, ie., such that $\widetilde{\mathsf{P}}\big(\widetilde{X}_n \le \widetilde{Y}_n\big)=1$ for all $n\ge0$.

Example 3.29. [Random-to-top shuffling] *For many practical purposes it is important to generate a random permutation of a collection of cards 1, 2, …, n. One way is to define a Markov chain $(X_n)_{n\ge0}$ in the space of all possible permutations, and run it until the stationary distribution is reached. One of the simplest algorithms—the* random-to-top shuffling*—is defined as follows: for a given state X_n of the chain (ie., given permutation), find jth card (with j taken uniformly at random) and put it at the top of the deck; repeat this step until the stationary distribution[30] is reached. As jth card can be chosen either* by value *or* by position, *we get two versions of the algorithm!*

To study the approach to stationarity, one couples two copies of the Markov chain, $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$, one starting from a fixed state, *and other starting* from equilibrium, *and shuffles until both configurations X_n and Y_n agree.*

Exercise 3.30.⊛ Let $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$ be random-to-top shuffling (RTTS) Markov chains, recall Example 3.29. Suppose that a random card in X_n (by position) is chosen, say of value j, and then put at the top to get X_{n+1}. In Y_n find the (randomly positioned) card with number j and also move it to the top to get Y_{n+1}. Let T be the first moment when both chains $(X_n)_{n\ge0}$ and $(Y_n)_{n\ge0}$ agree.

a) Show that the described algorithm provides a coupling[31] of the RTTS (by position) for $(X_n)_{n\ge0}$ and the RTTS (by value) for $(Y_n)_{n\ge0}$.

b) Describe the distribution of T.

[29] This example is important for computer science.
[30] show that the stationary distribution is uniform in the space of all orderings!
[31] try coupling RTTS (by value) and RTTS (by value) chains!