

3 Coupling

Two random variables, say X and Y , are *coupled* if they are defined on the same probability space. To couple two given variables X and Y means to define a random vector (\tilde{X}, \tilde{Y}) with joint probability $\tilde{P}(\cdot, \cdot)$ on some probability space¹⁴ so that the marginal distribution of \tilde{X} coincides with the distribution of X and the marginal distribution of \tilde{Y} coincides with the distribution of Y . Recall the following example:

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

		0	1	\tilde{X}
(A)	0	$q_1 q_2$	$q_1 p_2$	q_1
	1	$p_1 q_2$	$p_1 p_2$	p_1
	\tilde{Y}	q_2	p_2	

		0	1	\tilde{X}
(B)	0	q_2	$p_2 - p_1$	q_1
	1	0	p_1	p_1
	\tilde{Y}	q_2	p_2	

It is easy to see that in both cases¹⁵

$$\tilde{X} \sim \text{Ber}(p_1), \tilde{Y} \sim \text{Ber}(p_2),$$

though in the first case \tilde{X} and \tilde{Y} are independent, whereas in the second case we have $\tilde{P}(\tilde{X} \leq \tilde{Y}) = 1$.

3.1 Stochastic domination

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

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

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

¹⁴A 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!

¹⁵and in fact, every convex linear combination of these two tables provides a joint distribution with the same marginals.

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

$$\mathbb{P}(X > x) \leq \mathbb{P}(Y > x) \quad (3.1)$$

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

Remark 3.3. If $X \preceq Y$ and $g(\cdot) \geq 0$ is an arbitrary increasing function on \mathbb{R} , then $g(X) \preceq g(Y)$. If, in addition, $X \geq 0$, $Y \geq 0$, and $g(\cdot)$ is smooth with $g(0) = 0$, then

$$\mathbb{E}g(X) \equiv \int_0^\infty g'(z)\mathbb{P}(X > z) dz \leq \int_0^\infty g'(z)\mathbb{P}(Y > z) dz \equiv \mathbb{E}g(Y). \quad (3.2)$$

Example 3.4. If X is a random variable and $a \geq 0$ is a fixed constant, it is immediate that $Y = a + X$ stochastically dominates X , i.e., $X \preceq Y$: for each $x \in \mathbb{R}$ we have $\mathbb{P}(X > x) \leq \mathbb{P}(X > x - a) = \mathbb{P}(Y > x)$. Similarly, if $a \geq 1$ and $Z = aX \geq 0$, then $X \preceq Z$.

In the setup of Example 3.1, if $X \sim \text{Ber}(p_1)$ and $Y \sim \text{Ber}(p_2)$ then X is stochastically smaller than Y (i.e., $X \preceq Y$) if and only if $p_1 \leq p_2$; moreover, this is equivalent to existence of a coupling (\tilde{X}, \tilde{Y}) of X and Y in which these variables are ordered with probability one, $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$. The next result shows that this is a rather generic situation.

Lemma 3.5. A random variable X is stochastically smaller than a random variable Y if and only if there exists a coupling (\tilde{X}, \tilde{Y}) of X and Y such that $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$.

Remark 3.6. Notice that one claim of Lemma 3.5 is immediate from

$$\mathbb{P}(x < X) \equiv \tilde{\mathbb{P}}(x < \tilde{X}) = \tilde{\mathbb{P}}(x < \tilde{X} \leq \tilde{Y}) \leq \tilde{\mathbb{P}}(x < \tilde{Y}) \equiv \mathbb{P}(x < Y);$$

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

Example 3.7. If $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, p)$ with $m \leq n$, then $X \preceq Y$. Indeed, Let $Z_1 \sim \text{Bin}(m, p)$ and $Z_2 \sim \text{Bin}(n - m, p)$ be independent variables defined on the same probability space. We then put $\tilde{X} = Z_1$ and $\tilde{Y} = Z_1 + Z_2$ so that $\tilde{Y} - \tilde{X} = Z_2 \geq 0$ with probability one, $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$, and $X \sim \tilde{X}$, $Y \sim \tilde{Y}$.

Example 3.8. If $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\mu)$ with $\lambda \leq \mu$, then $X \preceq Y$. Indeed, Let $Z_1 \sim \text{Poi}(\lambda)$ and $Z_2 \sim \text{Poi}(\mu - \lambda)$ be independent variables defined on the same probability space.¹⁶ We then put $\tilde{X} = Z_1$ and $\tilde{Y} = Z_1 + Z_2$ so that $\tilde{Y} - \tilde{X} = Z_2 \geq 0$ with probability one, $\tilde{\mathbb{P}}(\tilde{X} \leq \tilde{Y}) = 1$, and $X \sim \tilde{X}$, $Y \sim \tilde{Y}$.

Example 3.9. Let $(X_n)_{n \geq 0}$ be a branching process with offspring distribution $\{p_m\}_{m \geq 0}$ and $X_0 = 1$. Let $(Y_n)_{n \geq 0}$ be a branching process with the same offspring distribution and $Y_0 = 2$. We can use stochastic domination to show that $\mathbb{P}(X_n = 0) \geq \mathbb{P}(Y_n = 0)$ for all n .

¹⁶here and below we assume that $Z \sim \text{Poi}(0)$ means that $\mathbb{P}(Z = 0) = 1$.

Indeed, it is enough to show that $X_n \preceq Y_n$ for all $n \geq 0$. To this end consider two independent branching processes, $(Z'_n)_{n \geq 0}$ and $(Z''_n)_{n \geq 0}$, having the same offspring distribution $\{p_m\}_{m \geq 0}$ and satisfying $Z'_0 = Z''_0 = 1$. We then put $\tilde{X}_n = Z'_n$ and $\tilde{Y}_n = Z'_n + Z''_n$ for all $n \geq 0$, so that $\tilde{Y}_n - \tilde{X}_n = Z''_n \geq 0$, i.e., $\tilde{P}(\tilde{X}_n \leq \tilde{Y}_n) = 1$ and $X_n \sim \tilde{X}_n$, $Y_n \sim \tilde{Y}_n$ for all $n \geq 0$.

Exercise 3.10. For a given offspring distribution $\{p_m\}_{m \geq 0}$, let $(X_n)_{n \geq 0}$ be the branching process with $X_0 = k$ and let $(Y_n)_{n \geq 0}$ be the branching process with $Y_0 = l$, $k < l$. Show that $X_n \preceq Y_n$ for all $n \geq 0$.

Exercise 3.11. Let X, Y, Z be random variables in \mathbb{Z}_+ . If $X \preceq Y$ and Z is independent of $\{X, Y\}$, show that $X + Z \preceq Y + Z$.

Let $\{X_i, Y_i\}_{i=1, \dots, n}$ be random variables in \mathbb{Z}_+ . If the pairs (X_i, Y_i) are mutually independent and $X_i \preceq Y_i$ for each i , show that $X_1 + \dots + X_n \preceq Y_1 + \dots + Y_n$.

Exercise 3.12. Let $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ be standard branching processes with $X_0 = Y_0 = 1$. Assume that the offspring distribution of X is stochastically smaller than that of Y ($X_1 \preceq Y_1$), i.e., for all integer $k \geq 0$, $P(X_1 > k | X_0 = 1) \leq P(Y_1 > k | Y_0 = 1)$. Show that $X_n \preceq Y_n$ for all $n \geq 0$.

3.2 Total variation distance

Definition 3.13. [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

$$d_{\text{TV}}(\mu, \nu) \stackrel{\text{def}}{=} \sup_A |\mu(A) - \nu(A)|. \quad (3.3)$$

If X, Y are random variables in \mathbb{Z}^+ with respective p.m.f.s $\{p\} = \{p_k\}_{k \geq 0}$ and $\{q\} = \{q_\ell\}_{\ell \geq 0}$, then

$$d_{\text{TV}}(X, Y) \stackrel{\text{def}}{=} d_{\text{TV}}(\{p\}, \{q\}) = \sup_{A \subseteq \mathbb{Z}^+} \left| \sum_{k \in A} p_k - \sum_{k \in A} q_k \right|.$$

In general if X, Y are two discrete random variables taking values in the same space \mathcal{X} , with $P(X = x) = p_x, P(Y = x) = q_x$ for $x \in \mathcal{X}$, then we define

$$d_{\text{TV}}(X, Y) \stackrel{\text{def}}{=} d_{\text{TV}}(\{p\}, \{q\}) = \sup_{A \subseteq \mathcal{X}} \left| \sum_{x \in A} p_x - \sum_{x \in A} q_x \right|.$$

Example 3.14. If $X \sim \text{Ber}(p_1)$ and $Y \sim \text{Ber}(p_2)$ we have

$$d_{\text{TV}}(X, Y) = \max\{|p_1 - p_2|, |q_1 - q_2|\} = |p_1 - p_2| = \frac{1}{2}(|p_1 - p_2| + |q_1 - q_2|).$$

Exercise 3.15. Let probability measures μ and ν have respective p.m.f.s $\{p_x\}$ and $\{q_y\}$. Show that the total variation distance between μ and ν is

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

Deduce that $d_{\text{TV}}(\cdot, \cdot)$ is a distance between probability measures¹⁷ (ie., it is non-negative, symmetric, and satisfies the triangle inequality) such that $d_{\text{TV}}(\cdot, \cdot) \leq 1$ for all probability measures μ and ν .

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

Example 3.16. [Maximal Coupling] Let random variables X and Y be such that $P(X = x) = p_x$ and $P(Y = y) = q_y$. Define the coupling (\tilde{X}, \tilde{Y}) via¹⁸

$$\hat{P}(\tilde{X} = \tilde{Y} = z) = \min(p_z, q_z), \quad (3.4)$$

$$\hat{P}(\tilde{X} = x, \tilde{Y} = y) = \frac{(p_x - \min(p_x, q_x))(q_y - \min(p_y, q_y))}{d_{\text{TV}}(\{p\}, \{q\})}, \quad x \neq y. \quad (3.5)$$

Exercise 3.17. Show that

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

and deduce that (3.4) is indeed a coupling of X and Y (ie., that (3.4) defines a probability distribution with correct marginals).

Example 3.18. Consider $X \sim \text{Ber}(p_1)$ and $Y \sim \text{Ber}(p_2)$ with $p_1 \leq p_2$. 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

$$\hat{P}(\tilde{X} \neq \tilde{Y}) = p_2 - p_1 = d_{\text{TV}}(X, Y).$$

Lemma 3.19. Let $\hat{P}(\cdot, \cdot)$ be the maximal coupling (3.4) of X and Y . Then for every other coupling $\tilde{P}(\cdot, \cdot)$ of X and Y we have

$$\tilde{P}(\tilde{X} \neq \tilde{Y}) \geq \hat{P}(\tilde{X} \neq \tilde{Y}) = d_{\text{TV}}(\tilde{X}, \tilde{Y}). \quad (3.6)$$

Proof. Summing the inequalities $\tilde{P}(\tilde{X} = \tilde{Y} = z) \leq \min(p_z, q_z)$ we deduce

$$\tilde{P}(\tilde{X} \neq \tilde{Y}) \geq 1 - \sum_z \min(p_z, q_z) = \sum_z (p_z - \min(p_z, q_z)) = d_{\text{TV}}(\{p\}, \{q\}),$$

in view of Exercise 3.17 and Example 3.18. □

¹⁷so that all probability measures form a metric space for this distance!

¹⁸By Exercise 3.15, if $d_{\text{TV}}(\{p\}, \{q\}) = 0$, we have $p_z = q_z$ for all z , and thus all off-diagonal terms in (3.4) vanish.

Remark 3.20. Notice that according to (3.6),

$$\tilde{\mathbb{P}}(\tilde{X} = \tilde{Y}) \leq \hat{\mathbb{P}}(\tilde{X} = \tilde{Y}) = 1 - d_{\text{TV}}(\tilde{X}, \tilde{Y}),$$

ie., the probability that $\tilde{X} = \tilde{Y}$ is maximised under the optimal coupling $\hat{\mathbb{P}}(\cdot, \cdot)$.

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

$$\hat{\mathbb{P}}(\tilde{X} = \tilde{Y} = 0) = 1 - p, \quad \hat{\mathbb{P}}(\tilde{X} = \tilde{Y} = 1) = pe^{-p},$$

and $\hat{\mathbb{P}}(\tilde{X} = \tilde{Y} = x) = 0$ for all $x > 1$. As a result,

$$d_{\text{TV}}(\tilde{X}, \tilde{Y}) \equiv \hat{\mathbb{P}}(\tilde{X} \neq \tilde{Y}) = 1 - \hat{\mathbb{P}}(\tilde{X} = \tilde{Y}) = p(1 - e^{-p}) \leq p^2. \quad (3.7)$$

Is either of the variables X and Y stochastically dominated by the other?

3.3 Applications to convergence

3.3.1 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 coupling¹⁹ of two random walks on a complete graph.

Example 3.22. On the complete graph K_m on m vertices, consider the random walk $(X_n)_{n \geq 0}$ with $X_0 = x$, such that at every step it jumps to any of the vertices of K_m uniformly at random (excluding the current one). To show that eventually $(X_n)_{n \geq 0}$ forgets its initial distribution, one couples $(X_n)_{n \geq 0}$ with another copy $(Y_n)_{n \geq 0}$, $Y_0 = y$, of this random walk, and shows that for large n both processes coincide with large probability.

To do this, we treat the pair $(\tilde{X}_n, \tilde{Y}_n)_{n \geq 0}$ 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

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

and set $\tilde{p}_{(x,y)(x',y')} = 0$ otherwise (ie., if $x = y$ and $x' \neq y'$). It is straightforward to check that this is a correct coupling.²⁰

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

Notice that the walks $(\tilde{X}_n)_{n \geq 0}$ and $(\tilde{Y}_n)_{n \geq 0}$ run independently until they meet, and from that moment onwards they move together. For $x \neq y$, denote

$$T \stackrel{\text{def}}{=} \min\{n \geq 0 : \tilde{X}_n = \tilde{Y}_n\}. \quad (3.8)$$

¹⁹Recall the first table in Example 3.1.

²⁰the argument for the other marginal is similar;

By a straightforward induction, it is easy to see that

$$\tilde{\mathbb{P}}(T > n) = \frac{1}{m} \left(1 - \frac{m-2}{(m-1)^2}\right)^n,$$

ie., T is (basically) a geometric random variable. Since \tilde{X}_n and \tilde{Y}_n coincide after they meet, and $\tilde{\mathbb{P}}, \tilde{X}_n, \tilde{Y}_n$ gives a coupling of (X_n, Y_n) for every n we have that

$$d_{\text{TV}}(X_n, Y_n) \leq \tilde{\mathbb{P}}(\tilde{X}_n \neq \tilde{Y}_n) = \tilde{\mathbb{P}}(T > n),$$

by ((3.6)). It therefore only remains to observe that the RHS above is exponentially small, $\tilde{\mathbb{P}}(T > n) \leq \left(1 - \frac{1}{m}\right)^n \leq e^{-n/m}$. In other words, the random walk $(X_n)_{n \geq 0}$ forgets its initial state $X_0 = x$ at least exponentially fast.

Notice that if $(Y_n)_{n \geq 0}$ 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 $|\mathbb{P}(X_n = j) - \frac{1}{m}| \leq e^{-n/m}$, ie., convergence towards the equilibrium distribution is at least exponentially fast.²¹

The next example is important for computer science.

Example 3.23. [Random walk on a hypercube] 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 $(W_k)_{k \geq 0}$ on \mathcal{H}_n is defined as follows: if $W_k = v \in \mathcal{H}_n$, select m , $1 \leq m \leq n$, uniformly at random and flip a fair coin. If the coin shows heads, set the m th coordinate of v to 1, otherwise set it to 0. The walk $(W_k)_{k \geq 0}$ is aperiodic and irreducible; as k increases, it tends to forget its initial state.

Exercise 3.24. Let $(W_k)_{k \geq 0}$ be the random walk on the hypercube \mathcal{H}_n , recall Example 3.23. Use Example 3.22 to study its distribution for large k .

Exercise 3.25. Let $(\xi_j)_{j \geq 1}$ be independent variables with $\mathbb{P}(\xi = 1) = 1 - \mathbb{P}(\xi = -1) = p_x$ and let $(\eta_j)_{j \geq 1}$ be independent variables with $\mathbb{P}(\eta = 1) = 1 - \mathbb{P}(\eta = -1) = p_y$. Define simple random walks $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ 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 $(X_n)_{n \geq 0}$ on \mathbb{Z} forgets its initial state x ; namely, for $p_x = p_y$ and $y - x = 2k$, construct a coupling of $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ similar to that of Example 3.22.

b) Show that the random walk $(X_n)_{n \geq 0}$ monotonously depends on p_x ; namely, for $x \leq y$ and $p_x \leq p_y$, use the ideas from Example 3.1 to construct a monotone coupling of $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$, ie., such that $\tilde{\mathbb{P}}(\tilde{X}_n \leq \tilde{Y}_n) = 1$ for all $n \geq 0$.

²¹?? gives a more precise information about this convergence.

Example 3.26. [Random-to-top shuffling] *For many practical purposes it is important to generate a random permutation of a collection of cards $1, 2, \dots, m$. One way is to define a Markov chain $(X_n)_{n \geq 0}$ in the space of all possible permutations (so at any time n , X_n will be a permutation of $\{1, \dots, m\}$), and run it until the stationary distribution is reached. The stationary distribution is as usual the uniform distribution on all possible permutations.*

One of the simplest algorithms—the random-to-top shuffling—is defined as follows: for a given state $X_n = (\sigma(1), \dots, \sigma(m))$ of the chain, find the j th card by position (with j taken uniformly at random between 1 and m) and put it at the top of the deck to get the new state $X_{n+1} = (\sigma(j), \dots, \sigma(j-1), \sigma(j+1), \dots)$. Note that we could take the j th card by value instead, and then the new state would be $(j, \dots, \sigma(k-1), \sigma(k+1), \dots)$ for the unique k such that $\sigma(k) = j$. Since j is chosen uniformly at random, this gives the same algorithm (a Markov chain with the same transition probabilities).

To study the approach to stationarity, one couples two copies of the Markov chain, $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$, one starting from a fixed state, and other starting from equilibrium, and shuffles until both configurations X_n and Y_n agree.

Exercise 3.27. *Let $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$ be random-to-top shuffling (RTTS) Markov chains, recall Example 3.26. 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 $T := \min\{k \geq 0 : Y_k = X_k\}$ be the coupling time for $(X_n)_{n \geq 0}$ and $(Y_n)_{n \geq 0}$.*

a) Show that the described algorithm provides a coupling²² of the RTTS (by position) for $(X_n)_{n \geq 0}$ and the RTTS (by value) for $(Y_n)_{n \geq 0}$.

b) Describe the distribution of T .

3.3.2 The Law of rare events

This subsection is optional and will not be examined.

The convergence result in Exercise 1.33 can also be derived using coupling:

Theorem 3.28. *Let $X = \sum_{k=1}^n X_k$, where $X_k \sim \text{Ber}(p_k)$ are independent random variables. Let, further, $Y \sim \text{Poi}(\lambda)$, where $\lambda = \sum_{k=1}^n p_k$. Then the maximal coupling of X and Y satisfies*

$$d_{\text{TV}}(\tilde{X}, \tilde{Y}) \equiv \hat{\mathbb{P}}(\tilde{X} \neq \tilde{Y}) \leq \sum_{k=1}^n (p_k)^2.$$

Proof. Write $Y = \sum_{k=1}^n Y_k$, where $Y_k \sim \text{Poi}(p_k)$ are independent rv's. Of course, $\mathbb{P}(\sum_{k=1}^n X_k \neq \sum_{k=1}^n Y_k) \leq \sum_{k=1}^n \mathbb{P}(X_k \neq Y_k)$ for every joint distribution of $(X_k)_{k=1}^n$ and $(Y_k)_{k=1}^n$. Let $\hat{\mathbb{P}}_k$ be the maximal coupling for the pair $\{X_k, Y_k\}$,

²²try coupling RTTS (by value) and RTTS (by value) chains!

and let \widehat{P}_0 be the maximal coupling for two sums. Notice that the LHS above is not smaller than $d_{TV}(\widetilde{X}, \widetilde{Y}) \equiv \widehat{P}_0(\widetilde{X} \neq \widetilde{Y})$; on the other hand, using the (independent) product measure $P = \widehat{P}_1 \times \cdots \times \widehat{P}_n$ on the right we deduce that then the RHS becomes just $\sum_{k=1}^n \widehat{P}_k(\widetilde{X}_k \neq \widetilde{Y}_k)$. The result now follows from (3.7). \square

Exercise 3.29. Let $X \sim \text{Bin}(n, \frac{\lambda}{n})$ and $Y \sim \text{Poi}(\lambda)$ for some $\lambda > 0$. Show that

$$\frac{1}{2} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| \leq d_{TV}(\widetilde{X}, \widetilde{Y}) \leq \frac{\lambda^2}{n} \quad \text{for every } k \geq 0. \quad (3.9)$$

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

Remark 3.30. Notice that (3.9) implies that if $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Poi}(np)$ then for every $k \geq 0$ the probabilities $\mathbb{P}(X = k)$ and $\mathbb{P}(Y = k)$ 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, i.e., they coincide in the first two decimal places.

3.4 Additional problems

Exercise 3.31. In the setting of Example 3.1 show that every convex linear combination of tables T_1 and T_2 , i.e., each table of the form $T_\alpha = \alpha T_1 + (1-\alpha)T_2$ with $\alpha \in [0, 1]$, gives an example of a coupling of $X \sim \text{Ber}(p_1)$ and $Y \sim \text{Ber}(p_2)$. Can you find all possible couplings for these variables?

Exercise 3.32. Generalise the inequality (3.2) to a broader class of functions $g(\cdot)$ and verify that if $X \preceq Y$, then $\mathbb{E}(X^{2k+1}) \leq \mathbb{E}(Y^{2k+1})$, $\mathbb{E}e^{tX} \leq \mathbb{E}e^{tY}$ with $t > 0$, $\mathbb{E}s^X \leq \mathbb{E}s^Y$ with $s > 1$ etc.

Exercise 3.33. Let $\xi \sim \mathcal{U}(0, 1)$ be a standard uniform random variable. For fixed $p \in (0, 1)$, define $X = 1_{\xi < p}$. Show that $X \sim \text{Ber}(p)$, a Bernoulli random variable with parameter p . Now suppose that $X = 1_{\xi < p_1}$ and $Y = 1_{\xi < p_2}$ for some $0 < p_1 \leq p_2 < 1$ and ξ as above. Show that $X \preceq Y$ and that $\mathbb{P}(X \leq Y) = 1$. Compare your construction to the second table in Example 3.1.

Exercise 3.34. Let $X \sim \text{Geom}(p)$ and $Y \sim \text{Geom}(r)$ be two geometric random variables. If $0 < p \leq r < 1$, are X and Y stochastically ordered? Justify your answer.

Exercise 3.35. Let $X \sim \Gamma(a, \lambda)$ and $Y \sim \Gamma(b, \lambda)$ be two gamma random variables.²³ If $0 < a \leq b < \infty$, are X and Y stochastically ordered? Justify your answer.

²³The 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 $\text{Exp}(\lambda)$.