

Coupling and total variational distance

Contents

1	Coupling	1
2	Examples	1
3	Comparing games	2
4	Enter coupling	2
5	Some R code	2
6	Total variational distance	4
6.1	Common formula	4
6.2	Coupling inequality	5
7	Examples	5
8	Summary	5
9	Version: 28 April 2021	5

1 Coupling

A **coupling** of two random variables X and Y is a pair of random variables (X', Y') with a joint distribution such that its marginal distributions are given by those of X and Y . Given two random variables X and Y , we can't do simple operations on them such as $X + Y$, unless we know their joint distribution, that is, if they even have one! If they have a joint distribution, then they live on the same probability space.

Often one can specify a coupling of X and Y , if we can find a function $F(X, U)$, where U is uniformly distributed on $[0, 1]$ and independent of X , so that $F(X, U)$ has the same law as Y . That is we can envision Y as a function of X and an additional randomization U .

2 Examples

- Independent coupling: this option is always available, but it is often *not* the most interesting or useful.
- Quantile coupling/Inverse transform: If F_X and F_Y are cdfs, and U is uniformly distributed, then $F_X^{-1}(U)$ and $F_Y^{-1}(U)$ are random variables with these respective cdfs.

- Thinning: If X is Bernoulli p and Z is an independent Bernoulli r , then XZ is Bernoulli pr . Thus (X, XZ) is a coupling of Bernoulli random variables with parameters p and pr , and $X \geq XZ$.

3 Comparing games

Suppose June plays a gambling game with her mum using $p = 2/3$ coin, betting on heads, winning 1 pound if the coin comes up heads, and losing 1 pound otherwise. We start with 0 pounds and allow the possibility of going negative. June stops playing as soon as she reaches 5 pounds. Suppose Tessa plays the same game with her dad using a $p = 1/2$ coin.

We can't say that June will be better off than Tessa for *sure*, since the coins are independent. Suppose we want to compare the expectation of the length of time W it takes for the game to end; it seems obvious that the mean time for June should be *less* than that of Tessa's. In fact, the mean time for Tessa is infinite.

4 Enter coupling

We can envision a coupling of these two games, where June is *always* better off.

- Let X_i be amount that June wins/loses on the i th flip, and Y_i is the corresponding amount for Tessa.
- We can arrange a coupling of X_i and Y_i so that $X'_i \geq Y'_i$, so that Tessa wins only when June wins. Use thinning!
- Under this coupling, we do have that $W'_J \leq W'_T$, so we can compute

$$\mathbb{E}W_J = \mathbb{E}W'_J \leq \mathbb{E}W'_T = \mathbb{E}W_T,$$

without knowing too much about W .

5 Some R code

Here, we provide some code to illustrates the two coupled games.

We first code and test the coupled coin-flips.

```
coupled<-function(){
  j= rbinom(1,1,2/3)
  t= rbinom(1,1,3/4)*j
  c(j,t)
}
S=replicate(10000, coupled())
mean(S[,1])
```

```
## [1] 0.6727
```

```
mean(S[,2])
```

```
## [1] 0.5063
```

Then, we run simulations to see long the game takes on average for June and Tessa.

```
exit<-function(p){
  x=0
  n=0
  while(x <5){
    x <- x+(2*rbinom(1,1,p) -1)
    n <- n+1
  }
  n
}

mean(replicate(1000, exit(2/3)))
```

```
## [1] 14.476
```

```
mean(replicate(25, exit(1/2)))
```

```
## [1] 1838.2
```

We could of run similar simulations without coupling. The point is, we established this result analytically using coupling. The code here is really just to illustrate the coupling. In the last bit of code, we note that in each simulation, under the coupling, Tessa never finishes before June.

```
exitcoupled<-function(){
  june =0
  tessa =0
  njune=0
  ntessa=0
  while(june < 5){
    v = coupled()
    june <- june + 2*v[1] -1
    tessa <- tessa + 2*v[2] -1
    njune <- 1+njune
    ntessa <- 1 + ntessa}
  while(tessa <5){
    tessa <- tessa+ 2*rbinom(1,1,0.5) -1
  }
  ntessa <- 1 + ntessa
}
c(njune, ntessa)
}
```

```
replicate(25, exitcoupled())
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
## [1,]    9   13    7   15   15    9   19    9    5
## [2,] 1491   95    9  503   77    9 4035   67   67
##      [,10] [,11] [,12] [,13] [,14] [,15] [,16]
## [1,]    9   29    7   11   13   31   17
## [2,]   43  125    7   25   15   81  119
##      [,17] [,18] [,19] [,20] [,21] [,22] [,23]
## [1,]   15   43    7    7   17    5   19
## [2,]   29 1055   37  107 30317   37   59
##      [,24] [,25]
## [1,]   19   15
## [2,]   21 2491
```

6 Total variational distance

Let X and Y be real-valued random variables. The **total variational** distance between X and Y is given by:

$$d_{TV}(X, Y) = 2 \cdot \sup_{A \text{ is an event}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

Note that the definition of $d_{TV}(X, Y)$ only depends on the laws of X and Y , individually; in particular, if we have that if (X', Y') is a coupling of X and Y , then

$$d_{TV}(X', Y') = d_{TV}(X, Y).$$

6.1 Common formula

Lemma 6.1. *Let X and Y be integer-valued random variables. We have that*

$$d_{TV}(X, Y) = \sum_{z \in \mathbb{Z}} |\mathbb{P}(X = z) - \mathbb{P}(Y = z)|.$$

The proof follows from the following simple fact: If

$$D = \{z \in \mathbb{Z} : \mathbb{P}(X = z) \geq \mathbb{P}(Y = z)\},$$

then

$$\sum_{z \in \mathbb{Z}} |\mathbb{P}(X = z) - \mathbb{P}(Y = z)| = \sum_{z \in D} (\mathbb{P}(X = z) - \mathbb{P}(Y = z)) + \sum_{z \in D^c} (\mathbb{P}(Y = z) - \mathbb{P}(X = z)).$$

Proof. Observe that for any $A \subset \mathbb{Z}$, we have

$$|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| = |\mathbb{P}(X \in A^c) - \mathbb{P}(Y \in A^c)|.$$

Thus using the set D , we see that

$$d_{TV}(X, Y) \geq \sum_{z \in \mathbb{Z}} |\mathbb{P}(X = z) - \mathbb{P}(Y = z)|.$$

For the other direction, note that

$$\begin{aligned} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| &= \left| \sum_{z \in A} \mathbb{P}(X = z) - \sum_{z \in A} \mathbb{P}(Y = z) \right| \\ &\leq \sum_{z \in A} |\mathbb{P}(X = z) - \mathbb{P}(Y = z)|. \end{aligned}$$

Thus it follows that

$$\begin{aligned} 2|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| &= |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| + |\mathbb{P}(X \in A^c) - \mathbb{P}(Y \in A^c)| \\ &\leq \sum_{z \in \mathbb{Z}} |\mathbb{P}(X = z) - \mathbb{P}(Y = z)|. \end{aligned}$$

□

Exercise 6.1. Let $X \sim \text{Bern}(p)$, $Y \sim \text{Bern}(q)$, and $W \sim \text{Poi}(p)$. Compute $d_{TV}(X, Y)$ and $d_{TV}(X, W)$.

6.2 Coupling inequality

Lemma 6.2. *If X and Y are jointly distributed random variables, then*

$$d_{TV}(X, Y) \leq 2\mathbb{P}(X \neq Y).$$

Proof. Let A be an event. Note that

$$\begin{aligned} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| &= \left| \mathbb{P}(X \in A, X = Y) + \mathbb{P}(X \in A, X \neq Y) - \mathbb{P}(Y \in A, X \neq Y) - \mathbb{P}(Y \in A, X = Y) \right| \\ &= |\mathbb{P}(X \in A, X \neq Y) - \mathbb{P}(Y \in A, X \neq Y)| \\ &\leq \mathbb{P}(X \neq Y). \end{aligned}$$

□

7 Examples

- Let X_1, \dots, X_n, X_{n+1} be i.i.d. Bernoulli with parameter p . Let S_k be their partial sum. We can bound

$$\begin{aligned} d_{TV}(S_n, S_{n+1}) &\leq 2\mathbb{P}(S_n \neq S_{n+1}) \\ &= 2\mathbb{P}(X_{n+1} = 1) \\ &= 2p \end{aligned}$$

- Let X be a six-sided fair die, and Y be a seven-sided fair die. Assume X and Y are independent. Let $X' = Y$, if $Y \neq 7$, otherwise, take $X' = X$. Clearly, X' has the same law as a six-sided fair die. Then

$$d_{TV}(X, Y) = d_{TV}(X', Y) \leq 2\mathbb{P}(X' \neq Y) = \frac{2}{7}.$$

8 Summary

We saw some basics of how coupling can be used as a tool in probability theory and in particular, how it can be used to bound the total variational distance between two random variables.

9 Version: 28 April 2021

- Rmd Source