CSC380: Principles of Data Science

Probability Primer

Xinchen Yu

Acknowledgement: Built on Jason Pacheco, Kwang-Sung Jun, Chicheng Zhang's slides

Annoucements

- Office hours will be out by end of this week
- Homework 1 out next Tuesday
- Readings before next Tuesday
  - Ch. 6 (WJ: Watkins, J., “An Introduction to the Science of Statistics: From Theory to Implementation”)

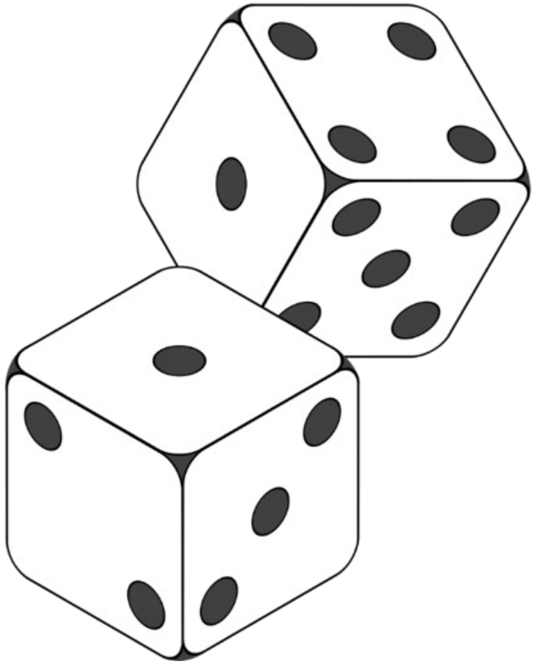
Outline

- Random Events and Probability
- Axioms of Probability
- Set Theory
- Law of Total Probability

Random Events and Probability

Random Events and Probability

Suppose we roll two fair dice…



- What are the possible outcomes?
- What is the *probability* of rolling **even** numbers?
- What is the *probability* of having two numbers sum to 6?
- If one die rolls 1, then what is the probability of the second die also rolling 1?

*…this is a **random process**.*

How to mathematically formulate outcomes and compute these probabilities?

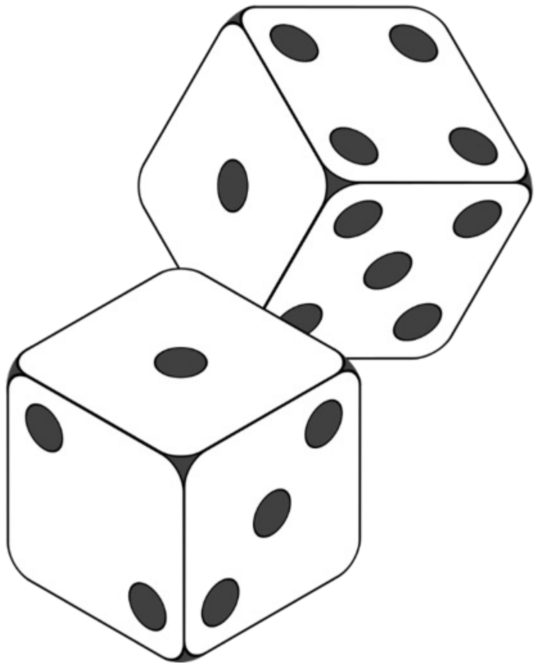
Random Events and Probability

Probability of a random event

$\approx$

Simulate the random process n times, the fraction of times this event happens

- How large should n be?
- Simulation results vary from trails?



Background: Numpy in Python

Numpy: numerical computing package

```
import numpy as np
np.random.randint(1,1+6,size=10)
=> array([5, 4, 1, 1, 1, 5, 5, 2, 4, 6])
```

`randint(low,high,size)`
: generate \`size\` random numbers in {low, low+1, …, high-1}

Numpy `array`

- Replaces python's **list** in numpy.
- More numerical functionality
- It's a 'vector' in mathematics.

```
a=np.array([1,2]); b=np.array([4,5])
a+b
⇒ np.array([5,7]) // elementwise addition
np.dot(a,b)
⇒ 14              // dot product
```

Random Events and Probability

Consider: What is the probability of having two numbers sum to 6?

```
import numpy as np
for n in [10,100,1_000,10_000,100_000]:
    res_dice1 = np.random.randint(1,6+1,size=n)
    res_dice2 = np.random.randint(1,6+1,size=n)
    res = [(res_dice1[i], res_dice2[i]) for i in range(len(res_dice1))]

    cnt = len(list(filter(lambda x: x[0] + x[1] == 6, res)))
    print("n=%6d, result: %.4f " % (n, cnt/n))
```

```
n=      10, result: 0.1000     n=      10, result: 0.1000
n=     100, result: 0.1200     n=     100, result: 0.1900
n=    1000, result: 0.1350     n=    1000, result: 0.1540
n=   10000, result: 0.1365     n=   10000, result: 0.1366
n=  100000, result: 0.1388     n=  100000, result: 0.1371
n= 1000000, result: 0.1385     n= 1000000, result: 0.1394
```

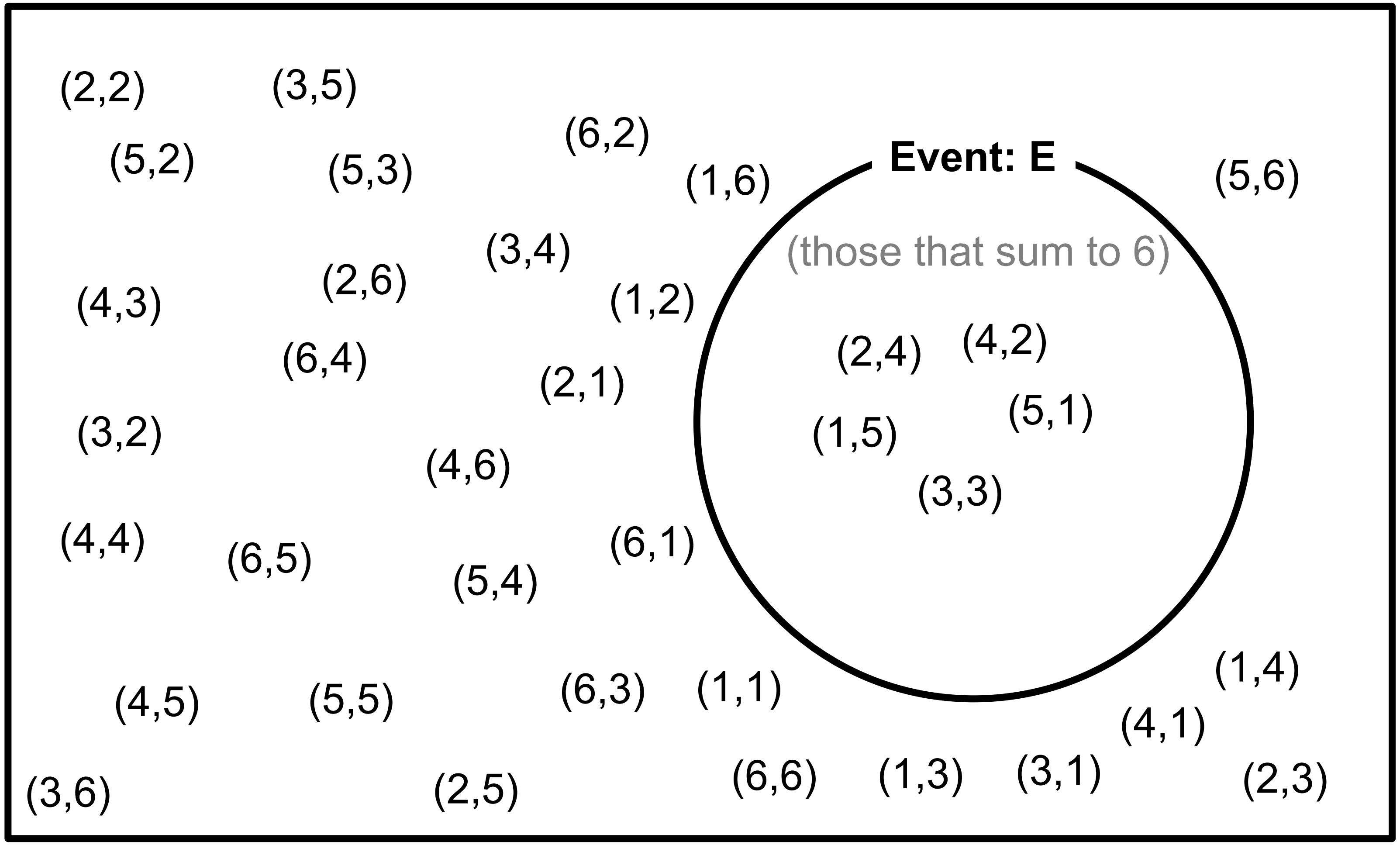
every time you run, you get a different result

however, the number seems to **converge** to 0.138-0.139

There seems to be a precise value that it will converge to.. what is it?

Random Events and Probability

Consider: What is the probability of having two numbers sum to 6?



Each outcome is equally likely.
by the **independence**
(will learn this concept later)
=> 1/36

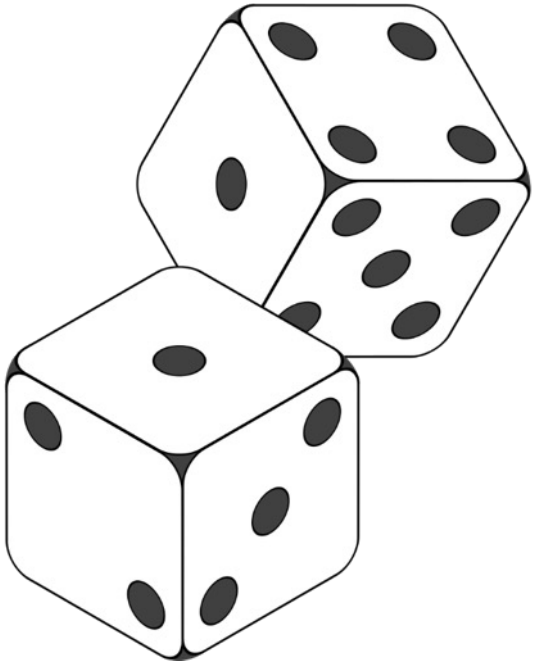
# of outcomes that sum to 6:
=> 5

answer:
(1/36) * 5 = 0.13888..

Random Events and Probability

- **Theoretical probability** describes how likely an event is going to occur based on math.

- **Experimental probability** describes how frequently an event actually occured in an experiment.



Mathematics of Probability

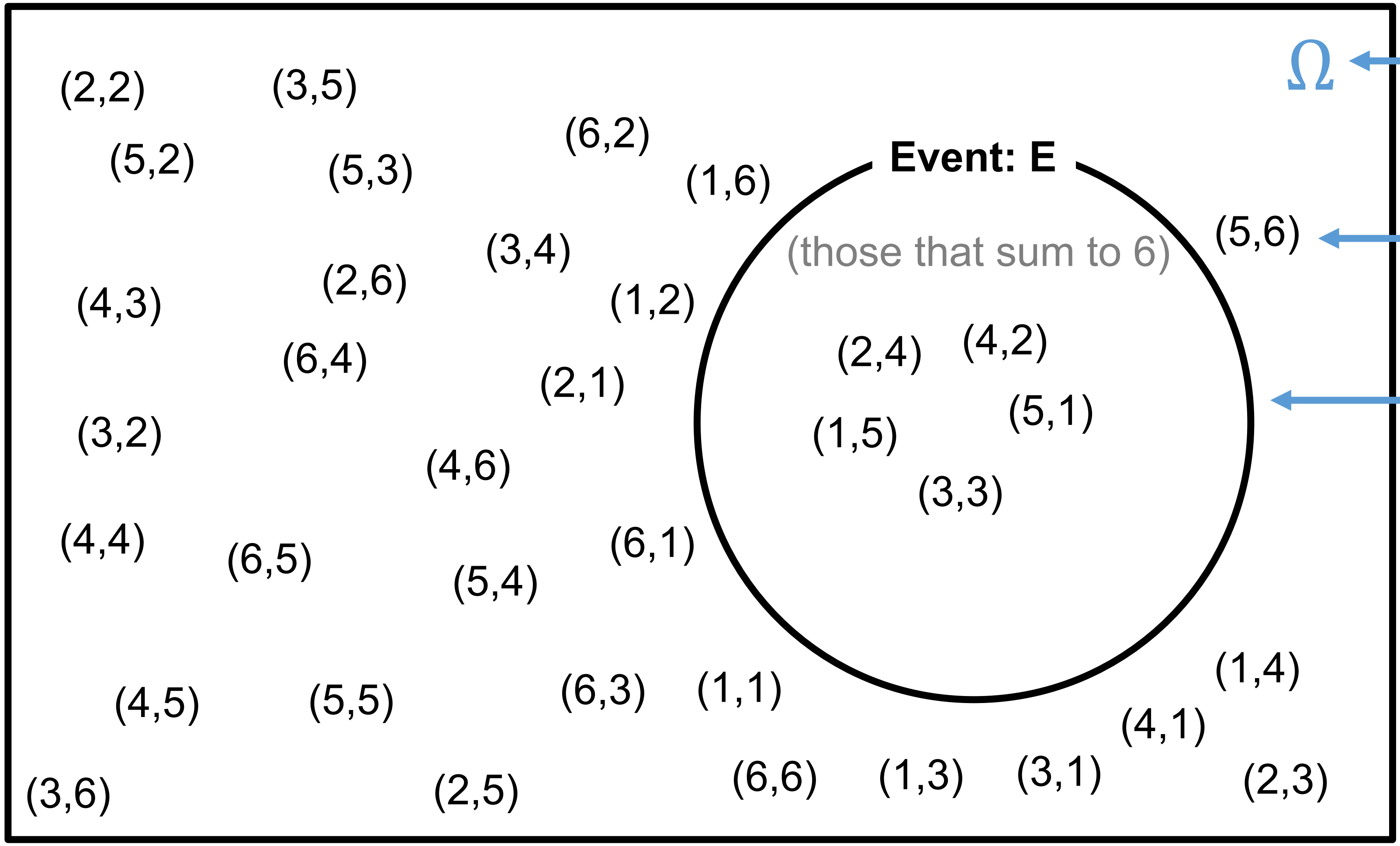
- **Probability** is a real-world phenomenon.

- But under what mathematical framework can we formulate **probability** so we can solve practical problems?
  - e.g., weather prediction, predicting the election outcome

- **Disclaimer**: not all mathematics correspond to real-world phenomenon (e.g., Banach–Tarski paradox). Fortunately, we will not talk about this in our lecture ☺

Random Events and Probability

Consider: What is the probability of having two numbers sum to 6?



Outcome space consists of all possible outcomes

Each **outcome** $\omega \in \Omega$.

Event is a collection of outcomes and a subset of the outcome space

$$E \subseteq \Omega$$

Note: outcome = sample

Random Events and Probability

Some examples of events…

Q: how many such pairs?

9

- Both even numbers

$$E^{\text{even}} = \{(2,2),(2,4),\ldots,(6,4),(6,6)\}$$

- The sum of both dice is even,

$$E^{\text{sum even}} = \{(1,1),(1,3),(1,5),\ldots,(2,2),(2,4),\ldots\}$$

- The sum is greater than 12,

$$E^{\text{sum}>12} = \emptyset$$

We can talk about impossible outcomes

Axioms of Probability

Random Events and Probability

But, what is probability, really?

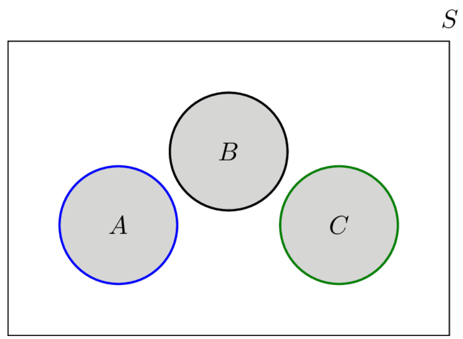
(e.g., can explain the probability of seeing an event when throwing two dice)

Mathematicians have found a set of conditions that 'makes sense'.

- Probability is a map P. ⇒ i.e., takes in an event, spits out a real value
- P must map events to a real value in interval [0,1].
- P is a (valid) **probability distribution** if it satisfies the following **axioms of probability**,

1. For any event E, $\mathrm{P}(\mathrm{E}) \geq 0$
2. $P(\Omega) = 1$
3. For any sequence of disjoint events $E_1, E_2, E_3, \ldots$

$$P\left(\bigcup_{i \geq 1} E_i\right) = \sum_{i \geq 1} P(E_i)$$



disjoint: intersection is empty

Random Events and Probability

- Many properties follows (i.e., can be proved mathematically)

$$\mathbb{P}(\emptyset) = 0$$

$$A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$$

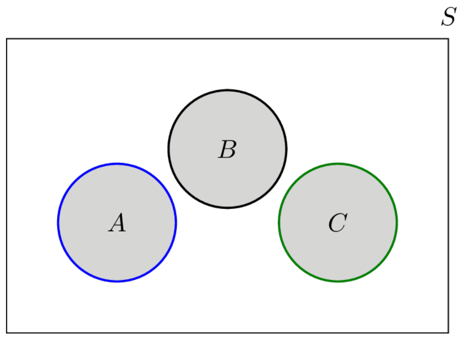
E.g., throw a die. A= getting 1, B=getting an odd number

$$0 \leq \mathbb{P}(A) \leq 1$$

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

$$A \bigcap B = \emptyset \implies \mathbb{P}\left(A \bigcup B\right) = \mathbb{P}(A) + \mathbb{P}(B).$$

E.g., A= getting 1, B=getting 3 or 5



(I recommend that you maintain your own version of cheat sheet!)

Random Events and Probability

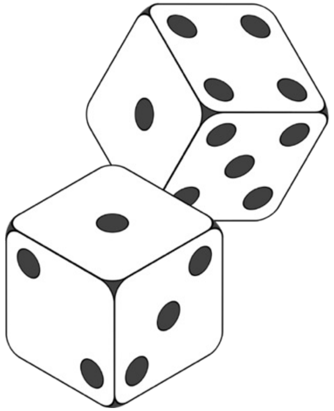
Special case

Assume each outcome is equally likely, and sample space is finite, then the probability of event is:

$$P(E) = \frac{|E|}{|\Omega|}$$

Number of elements in event set

Number of possible outcomes (36)



This is called uniform probability distribution

Q: What axiom we are using?
=> Axiom 3

(Fair) Dice Example: Probability that we roll even numbers,

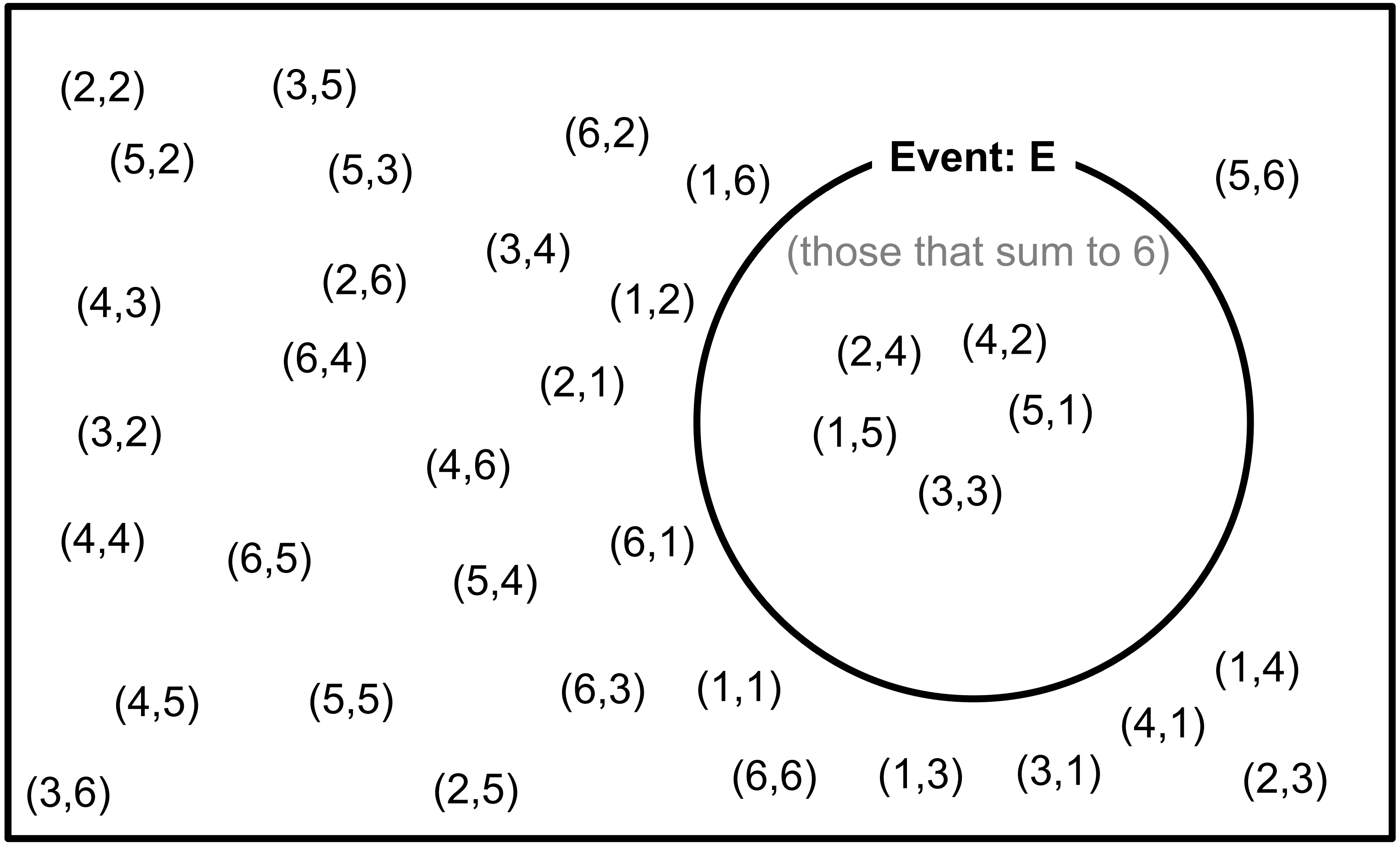
$$P((2,2) \cup (2,4) \cup \ldots \cup (6,6)) = P((2,2)) + P((2,4)) + \ldots + P((6,6))$$

$$= \frac{1}{36} + \frac{1}{36} + \ldots + \frac{1}{36} = \frac{9}{36}$$

9 Possible outcomes, each with equal probability of occurring

Random Events and Probability

Consider: What is the probability of having two numbers sum to 6?



Each outcome is equally likely.
by the **independence**
(will learn this concept later)
=> 1/36

\# of outcomes that sum to 6:
=> 5

answer:
(1/36) * 5 = 0.13888..

$$P(E) = \frac{|E|}{|\Omega|}$$

Set Theory

Set Theory

Two dice example: *Suppose*

E_1 : *First* die equals 1 E_2 : *Second* die equals 1

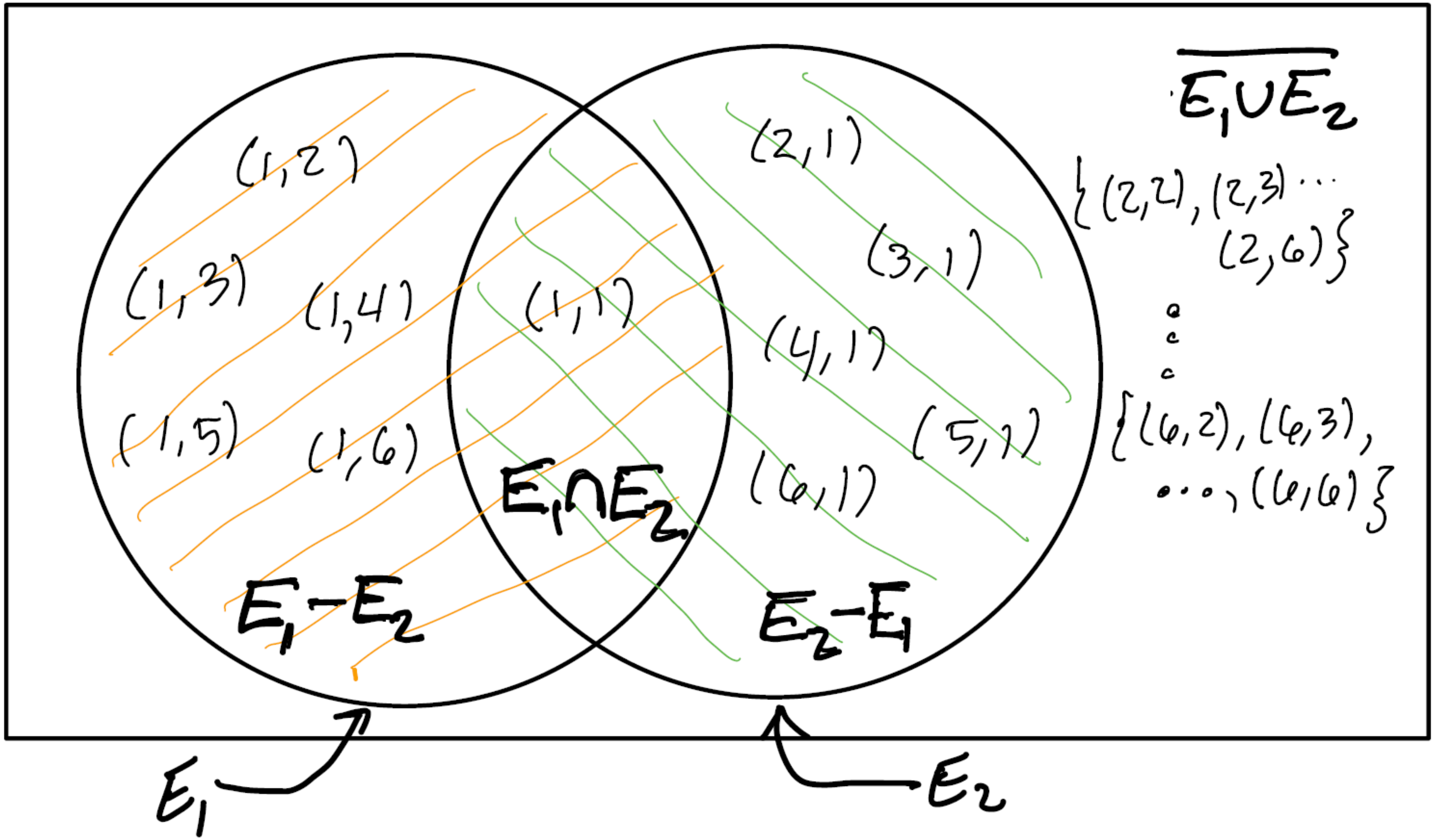
$$E_1 = \{(1,1),(1,2),\dots,(1,6)\} \qquad E_2 = \{(1,1),(2,1),\dots,(6,1)\}$$

Operators on events:

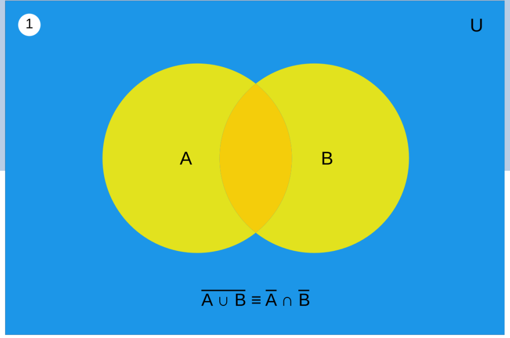
Operation	Value	Interpretation
$E_1 \cup E_2$	$\{(1,1),(1,2),\dots,(1,6),(2,1),\dots,(6,1)\}$	Any die rolls 1
$E_1 \cap E_2$	$\{(1,1)\}$	Both dice roll 1
$E_1 \setminus E_2$ $\boxed{(= E_1 - E_2 \coloneqq E_1 \cap E_2^c)}$	$\{(1,2),(1,3),(1,4),(1,5),(1,6)\}$	Only the first die rolls 1
$\overline{E_1 \cup E_2}$ $\boxed{(= (E_1 \cup E_2)^c)}$	$\{(2,2),(2,3),\dots,(2,6),(3,2),\dots,(6,6)\}$	No die rolls 1

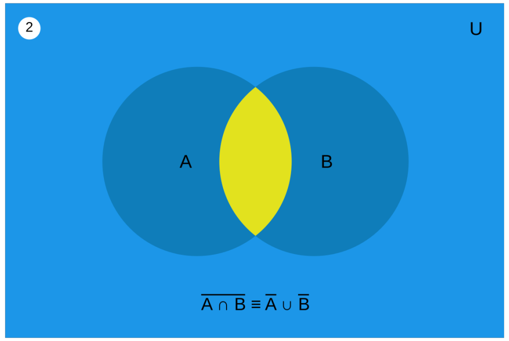
Set Theory

Can interpret these operations as a Venn diagram…



Set Theory





More results

- $\neg(\bigcup_n A_n) = \bigcap_n \neg A_n$, $\neg(\bigcap_n A_n) = \bigcup_n \neg A_n$ DEMORGAN

  Special case: $\neg(A \cup B) = \neg A \cap \neg B$

  Notation: $\neg A := A^c$

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$. // distributive law

  $A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$, $A \cup (\cap_i B_i) = \cap_i (A \cup B_i)$

- $B = \Omega \cap B = (A \cup \neg A) \cap B = (A \cap B) \cup (\neg A \cap B)$ // by distributive law

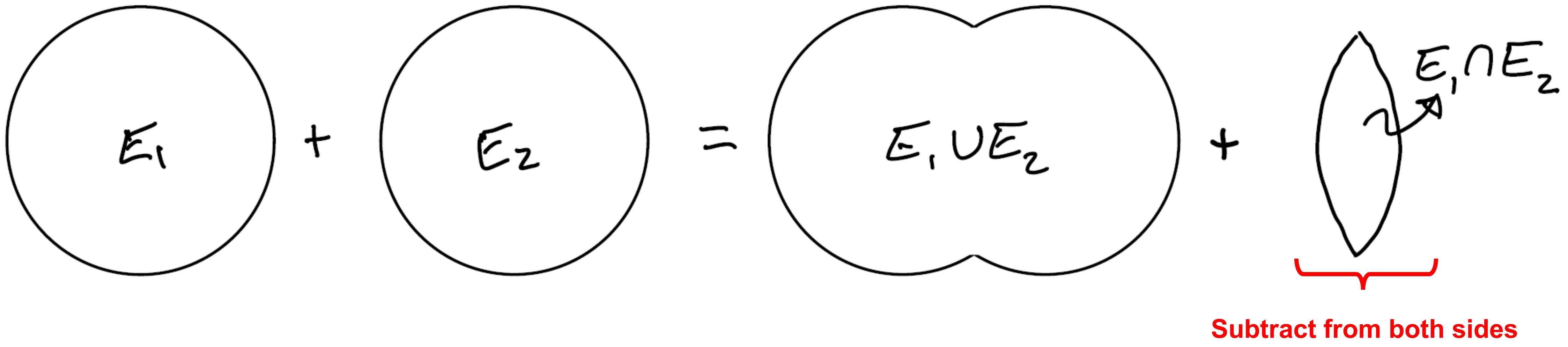
TIP: always draw a picture to visualize these identities!

For more, see https://math.libretexts.org/Courses/Mount_Royal_University/MATH_1150%3A_Mathematical_Reasoning/2%3A_Basic_Concepts_of_Sets/2.3%3A_Properties_of_Sets

Lemma: (inclusion-exclusion rule) *For <u>any</u> two events* E_1 *and* E_2,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Graphical Proof:

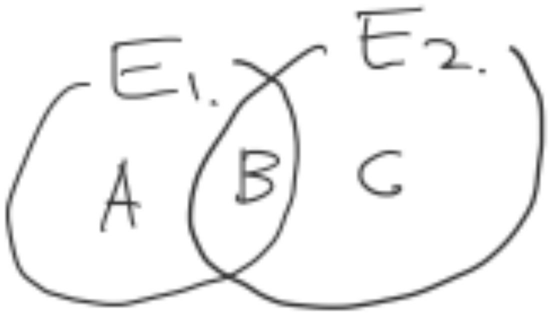


Alternative Proof

Lemma: *For <u>any</u> two events* E_1 *and* E_2 *,*

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Alternative proof:



$A = E_1 - (E_1 \cap E_2)$

$B = E_1 \cap E_2$

$C = E_2 - (E_1 \cap E_2)$.

$$
\begin{aligned}
&P(E_1 \cup E_2) \\
&= P(A \cup B \cup C) \\
&= P(A) + P(B) + P(C) && \text{(by axiom 3)} \\
&= P(A) + P(B) + P(B) + P(C) - P(B) \\
&= P(A \cup B) + P(B \cup C) - P(B) && \text{(by axiom 3)}
\end{aligned}
$$

Random Events and Probability

Exercise:

Quiz candidate

- Consider rolling two fair dice
- E_1: two dice sum to 6
- E_2: second die is even
- Compute the numerical value of $P(E_1 \cup E_2)$. Hint: Use inclusion-exclusion rule.

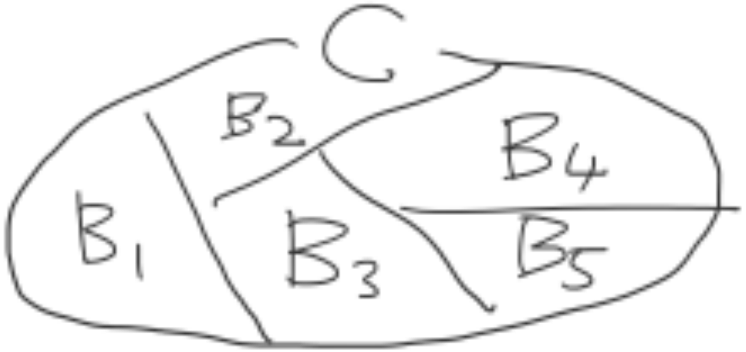
$P(E_1) = 5/36$
$P(E_2) = 12/36$
$P(E_1 \cap E_2) = 2/36$

answer: 21/36

Law of Total Probability

Random Events and Probability

[Def] The set of events $\{B_i\}_{i=1}^{n}$ **partitions** outcome space $C \Leftrightarrow \cup_i B_i = C$ and $B_1, B_2, \ldots$ are disjoint.



<u>Law of total probability</u>*:* Let A be an event. For events $\mathrm{B}_1, \mathrm{B}_2, \ldots$ that partitions Ω, we have

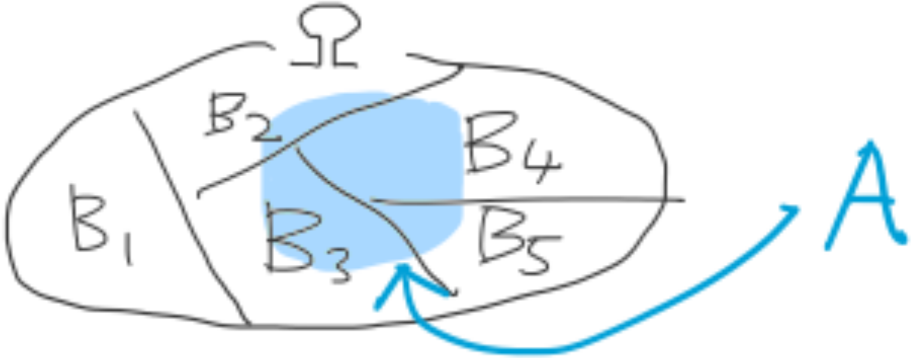
$$P(A) = \sum_i P(A \cap B_i)$$

Now, $\{A \cap B_i\}_{i=1}^{n}$ partitions A

Q: Why is this true?

A: Axiom 3!

$$A = A \cap \Omega = A \cap (\cup_i B_i) = \cup_i (A \cap B_i)$$



Random Events and Probability

<u>Law of total probability</u>: Let A be an event. For any events $\mathrm{B}_1, \mathrm{B}_2, \ldots$ that partitions Ω, we have

$$P(A) = \sum_i P(A \cap B_i)$$

Example Roll two fair dice. Let X be the <u>outcome of the first die</u>. Let Y be the <u>sum of both dice</u>. What is the probability that both dice sum to 6 (i.e., Y=6)?

quiz candidate

$$\mathrm{P}(A, B) := P(A \cap B)$$

$$\begin{aligned} p(Y=6) &= \sum_{x=1}^{6} p(Y=6, X=x) \\ &= p(Y=6, X=1) + p(Y=6, X=2) + \ldots + p(Y=6, X=6) \\ &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + 0 = \frac{5}{36} \end{aligned}$$

Summary So Far

- Most of the rules we learned is basically
  set theory + axiom 3

So, here is a generic workflow for computing P(A).

1. Use set theory and slice and dice A into a manageable partition of A where P(each piece of partition) is easy to compute.
2. Apply Axiom 3.

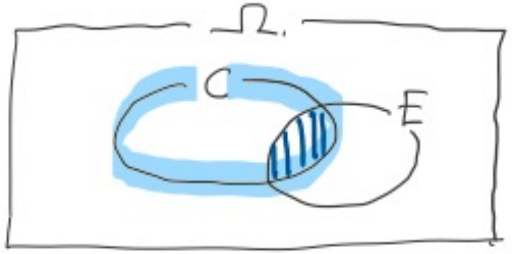
Conditional Probability

- Two fair dice example:
  - Suppose I roll two dice secretly and tell you that <u>one of the dice is 2</u>. C
  - **<u>In this situation</u>**, find the probability of two dice summing to 6. E

```
import numpy as np
for n in [10,100,1000,10_000,100_000, 1_000_000]:
    res_dice1 = np.random.randint(6,size=n) + 1
    res_dice2 = np.random.randint(6,size=n) + 1
    res = [(res_dice1[i], res_dice2[i]) for i in range(len(res_dice1))]

    conditioned = list(filter(lambda x: x[0] == 2 or x[1] == 2, res))
    n_eff = len(conditioned)

    cnt = len(list(filter(lambda x: x[0] + x[1] == 6, conditioned)))
    print("n=%9d, n_eff=%9d, result: %.4f " % (n, n_eff, cnt/n_eff))
```



compare:
without conditioning,
it was 0.138..

```
n=       10, n_eff=        4, result: 0.0000
n=      100, n_eff=       32, result: 0.2500
n=     1000, n_eff=      300, result: 0.1733
n=    10000, n_eff=     3002, result: 0.1742
n=   100000, n_eff=    30590, result: 0.1823
n=  1000000, n_eff=   305616, result: 0.1818
```

```
n=       10, n_eff=        3, result: 0.3333
n=      100, n_eff=       32, result: 0.0625
n=     1000, n_eff=      343, result: 0.2245
n=    10000, n_eff=     3062, result: 0.1897
n=   100000, n_eff=    30651, result: 0.1811
n=  1000000, n_eff=   305580, result: 0.1808
```