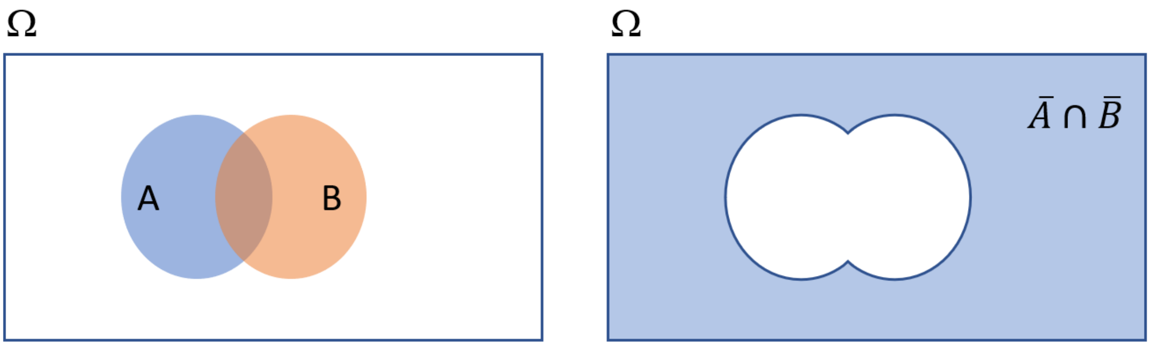
Exercise Set 2 - Solutions

Goals

1) Exercise basic probability operations
2) Understand the standard deviation and variance, and their difference
3) Familiarize yourself with random variables and their plots in Python

1 Warm-up on events operation

a) if A and B are independent events, show that $\bar{A}$ and $\bar{B}$ (the complementary events) are also independent.



Two events, A and B, are stochastically independent if and only if $P(A \cap B) = P(A) * P(B)$. We need to show that $P(\bar{A} \cap \bar{B}) = P(\bar{A}) * P(\bar{B})$ to prove independence.

Drawing the Venn diagram illustrates that $P(\bar{A} \cap \bar{B}) = 1 - P(A) - P(B) + P(A \cap B) = 1 - P(A) - P(B) + P(A) * P(B)$, as A and B are independent.

$1 - P(A) - P(B) + P(A) * P(B) = (1 - P(A)) * (1 - P(B)) = P(\bar{A}) * P(\bar{B})$.

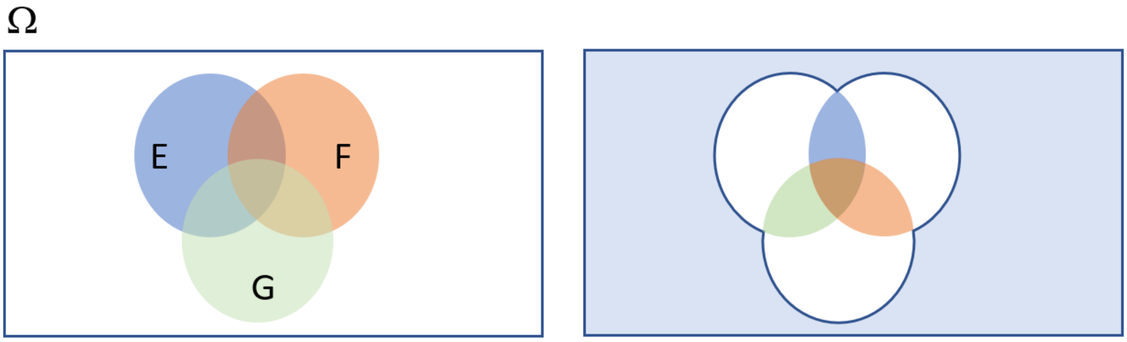
Therefore, also the complementary events of independent events are themselves independent.

b) Let E, F and G be three events. Find the expressions in terms of the probabilities P(E),P(F),P(G) for:

- "none of them is realized"

$P(\Omega \setminus E \setminus F \setminus G) = 1 - P(E) - P(F) - P(G) + P(E \cap F) + P(F \cap G) + P(E \cap G) - P(E \cap F \cap G)$

- "at maximum one is realized"



$1 - P(E \cap F) - P(F \cap G) - P(E \cap G) + 2P(E \cap F \cap G)$

- "exactly two are realized"

$P(E \cap F) + P(F \cap G) + P(E \cap G) - 3P(E \cap F \cap G)$

- "E and F are realized but not G"

$P(E \cap F) - P(E \cap F \cap G)$

2 Your odds at Roulette

a)

1) There are a total of 37 numbers possible. The odds to find a single number are thus $P = \frac{1}{37} \sim 2.7\%$.

2) The zero counts as colorless, the other 36 are half black, half red. Hence the Laplacian probability is given by $P = \frac{18}{37} \sim 48.6\%$.

3) $P = \frac{12}{37} \sim 32.4\%$.

4) $P = \frac{18}{37} \sim 48.6\%$. (zero counts as neither even nor odd)

b) We have to show that $P(Even \cap Red) = P(Even) * P(Red)$. We know from 2) and 4) that $P(Even) * P(Red) = \frac{324}{1369} \sim 23.7\%$. We also need to find the intersection, meaning we need to count all numbers that are both even and red at the same time numbers. A look at the list gives 8 even red numbers. Thus, $P(Even \cap Red) = \frac{8}{37} \sim 21.6\%$.

Hence, the events "red" and "even" are *not* independent. This means that if a roulette ball has been thrown, and your friend knows the number but only tells you the color that has been realized, you have a slightly higher chance than 50% to guess right because the events are dependent.

c)-e) See E2.ipnb Jupyter Noteb Python file on Moodle.

c) The key part of the code is:

```
r = np.random.randint(37)
print(r)
if r!=0:
```

```
    print("Pair!") if r%2==0 else print("Impair!")
    print("Rouge!") if r in [32,19,21,25,34,27,36,30,23,5,16,1,14,9,18,7,12,3] else print("Noir!
    print("Manque!") if r<19 else print ("Passe!")
    print("1st dozen!") if r<13 else (print("2nd dozen") if r<25  else print ("3rd dozen"))
```

There are of course other ways of realizing this, this is one of the most compact.

d) The key part of the code is:

```
c = 10
cList =[]
while c > 0:
    cList.append(c)
    r = np.random.randint(37)
    c = c+1 if r%2==1 else c-1
plt.plot(cList)
```

e) The key part of the code is:

```
iList = []
for j in range(1000):
    c = 10
    i = 0
    while c > 0:
        i = i+1
        r = np.random.randint(37)
        c = c+1 if r%2==1 else c-1
    iList.append(i)
plt.hist(iList)
plt.show()

iArray = np.array(iList)
print(np.mean(iArray))
print(np.median(iArray))
print(np.std(iArray)))
```

Clearly this distribution is very asymmetric and has some extreme outliers: Most of the games last between 10 (that is actually the minimum number, and only occurs if you lose every time) and 200 rounds, but occasionally they can last more than 1000 times. Therefore the mean does not represent the distribution well (in fact the mean and median are very different for example). The reason for this asymmetry is that there is always a certain probability (but smaller and smaller) probability that you are lucky and can play more and more rounds, so there is no "hard maximum", but there is a "hard minimum."

3 Modeling a random coin game experiment

In this game, we throw a coin four times and record the results.

a) Each coin throw has 2 possible outcomes, "heads" or "tails". We assign numbers to these cases, "heads"=0 and "tails"=1. Four consecutive coin throws form an elementary event in this game, denoted by a vector (coin1,coin2,coin3,coin4). For example, $(1, 0, 0, 1)$ denotes the elementary case "first coin tails, second and third coin heads, fourth coin tails".

b) There are 2 outcomes per coin, and 4 coin throws. Therefore Ω contains $2^4 = 16$ elementary events.

c)

- "we get heads at the second and fourth throw"

$\{(0,0,0,0),(1,0,0,0),(0,0,1,0),(1,0,1,0)\}$

- "we get heads only at the second and fourth throw"

$\{(1,0,1,0)\}$

- "we get at least one heads out of the four throws"

$\Omega \setminus \{(1,1,1,1)\}$

- "we get at least two heads out of the four throws"

$\{(1,1,0,0),(1,0,1,0),(1,0,0,1),(0,1,1,0),(0,1,0,1),$
$(0,0,1,1),(0,0,0,1),(0,0,1,0),(0,1,0,0),(1,0,0,0),(0,0,0,0)\}$

d) This can be computed by counting the number n of cases in each event in c). The total number of events is $2^4 = 16$. Therefore, the probability of each event is $\frac{n}{16}$.

e) Even though the family composition and coin throws are very different processes, within the over-simplified assumption used here they share the same mathematical description: independent random processes with a probability of $p = 0.5$. Hence the results from c),d) are just the same, mapping "heads/tails" to "girl/boy".

The reality in the latter case is much more complex of course, for biological and social reasons, and even coin flips are not exactly 50:50 - there is a slightly higher chance for the coin to fall onto the same orientations that it started, see https://doi.org/10.48550/arXiv.2310.04153 .

f) For the hair colour of two children, the approximation that the probabilities are uncorrelated is of course a very bad assumption: If one child has blonde hair, the probability of another child having blone hair is much higher than it would be in general.

4 Revisit steel cable data (Exercise Set 1)

The variance is 2.79MPa^2, and the standard deviation is 1.67MPa. The mean is 11.31MPa, thus the interval in the exercise is [9.64,12.98]MPa. Counting gives a total of 14 out of 20 points in this interval, a fraction $p = \frac{14}{20} = 70\%$.