Math 461 Spring 2024

Renming Song

University of Illinois Urbana-Champaign

January 22, 2024

Outline

Outline

I have setup a HW1 folder in the Moodle page. Please submit your HW1 in ONE pdf file via that folder. Make sure the quality of your file is good enough. The deadline for submitting HW1 is next Friday, 01/26, before the end of our lecture.

Outline

There are lots of phenomena in nature whose outcome cannot be predicted with certainty in advance, but the set of all the possible outcomes is known. For instance, when you toss a coin, you do not know whether “Heads” or “Tails” will appear, but you do know the outcome will be either ‘Heads” or “Tails”. These are what we call random phenomena or random experiments. Probability theory is concerned with such random experiments.

Consider a random experiment. The set of all the possible outcomes is called the *sample space* of the experiment. We usually denote the sample space by S.

There are lots of phenomena in nature whose outcome cannot be predicted with certainty in advance, but the set of all the possible outcomes is known. For instance, when you toss a coin, you do not know whether “Heads” or “Tails” will appear, but you do know the outcome will be either ‘Heads” or “Tails”. These are what we call random phenomena or random experiments. Probability theory is concerned with such random experiments.

Consider a random experiment. The set of all the possible outcomes is called the *sample space* of the experiment. We usually denote the sample space by S.

Tossing a coin. $S = \{H, T\}$. H stands for "Heads", and T stands for "Tails".

Tossing a (6-sided) die. $S = \{1, 2, 3, 4, 5, 6\}$.

Tossing a coin twice: $S = \{HH, HT, TH, TT\}$.

Tossing a (6-sided) die twice. $S = \{(i, j) : i, j = 1, \ldots, 6\}$.

Tossing a coin. $S = \{H, T\}$. H stands for "Heads", and T stands for "Tails".

Tossing a (6-sided) die. $S = \{1, 2, 3, 4, 5, 6\}$.

Tossing a coin twice: $S = \{HH, HT, TH, TT\}$.

Tossing a (6-sided) die twice. $S = \{(i, j) : i, j = 1, \ldots, 6\}$.

Tossing a coin. $S = \{H, T\}$. H stands for “Heads”, and T stands for “Tails”.

Tossing a (6-sided) die. $S = \{1, 2, 3, 4, 5, 6\}$.

Tossing a coin twice: $S = \{HH, HT, TH, TT\}$.

Tossing a (6-sided) die twice. $S = \{(i, j) : i, j = 1, \ldots, 6\}$.

Tossing a coin. $S = \{H, T\}$. H stands for “Heads”, and T stands for “Tails”.

Tossing a (6-sided) die. $S = \{1, 2, 3, 4, 5, 6\}$.

Tossing a coin twice: $S = \{HH, HT, TH, TT\}$.

Tossing a (6-sided) die twice. $S = \{(i, j) : i, j = 1, \ldots, 6\}$.

Keeping on tossing a coin until an H appears.
$S = \{H, TH, TTH, TTTH, \dots\}$.

Measuring the lifetime of a light-bulb. $S = [0, \infty)$.

Any subset E of the sample space S is known as an *event*. Some examples are

Keeping on tossing a coin until an H appears.
$S = \{H, TH, TTH, TTTH, \dots\}$.

Measuring the lifetime of a light-bulb. $S = [0, \infty)$.

Any subset E of the sample space S is known as an *event*. Some examples are

Keeping on tossing a coin until an H appears.
$S = \{H, TH, TTH, TTTH, \dots\}$.

Measuring the lifetime of a light-bulb. $S = [0, \infty)$.

Any subset E of the sample space S is known as an *event*. Some examples are

Tossing a coin. $E = \{H\}$.

Tossing a (6-sided) die. $E = \{2, 4, 6\}$.

Tossing a coin twice: $E = \{HH, HT\}$.

Tossing a (6-sided) die twice. $E =$ “the sum is 6”.

Tossing a coin. $E = \{H\}$.

Tossing a (6-sided) die. $E = \{2, 4, 6\}$.

Tossing a coin twice: $E = \{HH, HT\}$.

Tossing a (6-sided) die twice. $E =$ "the sum is 6".

Tossing a coin. $E = \{H\}$.

Tossing a (6-sided) die. $E = \{2, 4, 6\}$.

Tossing a coin twice: $E = \{HH, HT\}$.

Tossing a (6-sided) die twice. $E =$ "the sum is 6".

Tossing a coin. $E = \{H\}$.

Tossing a (6-sided) die. $E = \{2, 4, 6\}$.

Tossing a coin twice: $E = \{HH, HT\}$.

Tossing a (6-sided) die twice. $E =$ “the sum is 6”.

Keeping on tossing a coin until an H appears.
$E = \{H, TH, TTH, TTTH\}$.

Measuring the lifetime of a light-bulb. $E = [90, \infty)$.

We say that an event E occurs if the outcome of the experiment belongs to E.

Keeping on tossing a coin until an H appears.
$E = \{H, TH, TTH, TTTH\}$.

Measuring the lifetime of a light-bulb. $E = [90, \infty)$.

We say that an event E occurs if the outcome of the experiment belongs to E.

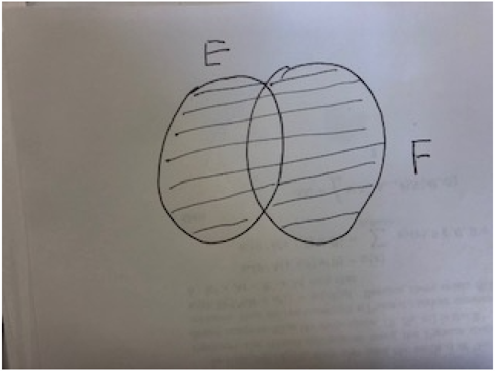
Keeping on tossing a coin until an H appears.
$E = \{H, TH, TTH, TTTH\}$.

Measuring the lifetime of a light-bulb. $E = [90, \infty)$.

We say that an event E occurs if the outcome of the experiment belongs to E.

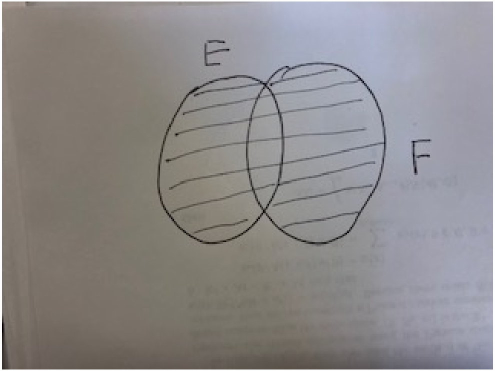
Events are simply subsets of the sample space, so we can talk about various set theoretical operations of events.

Union: $E \cup F$ occurs if and only if E or F occurs.



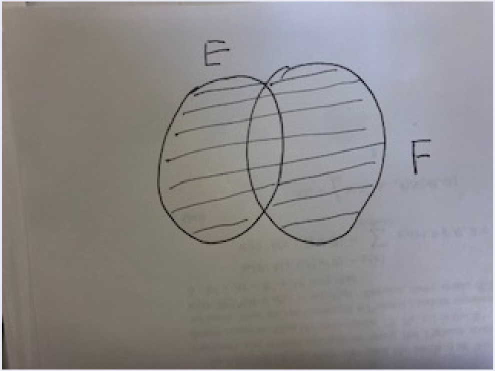
Events are simply subsets of the sample space, so we can talk about various set theoretical operations of events.

Union: $E \cup F$ occurs if and only if E or F occurs.

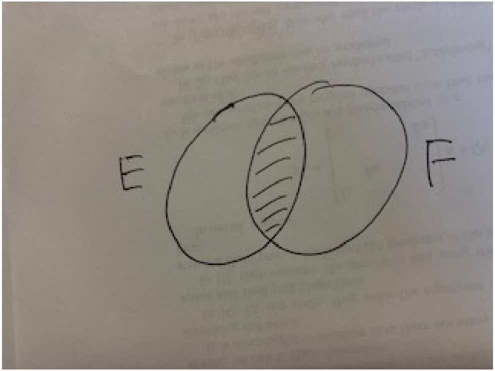


Events are simply subsets of the sample space, so we can talk about various set theoretical operations of events.

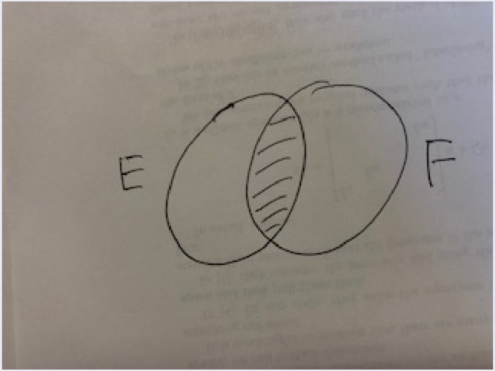
Union: $E \cup F$ occurs if and only if E or F occurs.



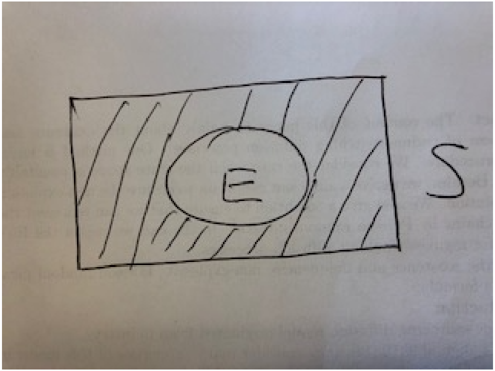
Intersection: $E \cap F$ occurs if and only if both E and F occur



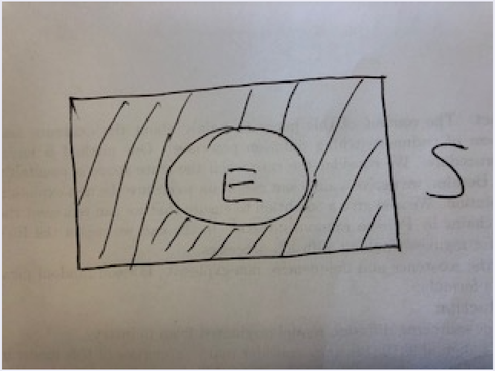
Intersection: $E \cap F$ occurs if and only if both E and F occur



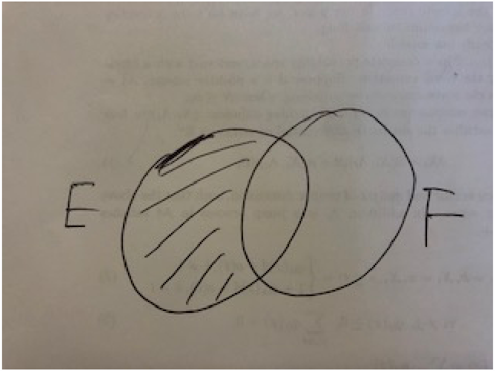
The complement of E, denoted as E^c, consists of all the elements of S which are not in E.



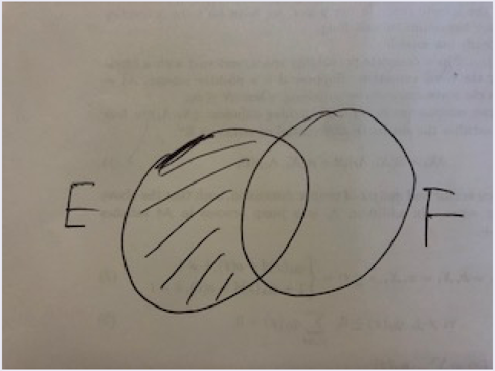
The complement of E, denoted as E^c, consists of all the elements of S which are not in E.



$E \setminus F = E \cap F^c$ consists of elements which are in E but not in F.



$E \setminus F = E \cap F^c$ consists of elements which are in E but not in F.



$E \subset F$ if every element of E is an element of F.

If $E \cap F = \emptyset$, then we say that E and F are disjoint, or mutually exclusive.

Similarly, we can define the union and intersection of more than 2 events

$$\cup_{i=1}^{n} E_i, \quad \cup_{i=1}^{\infty} E_i$$

and

$$\cap_{i=1}^{n} E_i, \quad \cap_{i=1}^{\infty} E_i.$$

$E \subset F$ if every element of E is an element of F.

If $E \cap F = \emptyset$, then we say that E and F are disjoint, or mutually exclusive.

Similarly, we can define the union and intersection of more than 2 events

$$\cup_{i=1}^{n} E_i, \quad \cup_{i=1}^{\infty} E_i$$

and

$$\cap_{i=1}^{n} E_i, \quad \cap_{i=1}^{\infty} E_i.$$

$E \subset F$ if every element of E is an element of F.

If $E \cap F = \emptyset$, then we say that E and F are disjoint, or mutually exclusive.

Similarly, we can define the union and intersection of more than 2 events

$$\cup_{i=1}^{n} E_i, \quad \cup_{i=1}^{\infty} E_i$$

and

$$\cap_{i=1}^{n} E_i, \quad \cap_{i=1}^{\infty} E_i.$$

Properties of set theoretical operations

Commutativity: $E \cup F = F \cup E$ and $E \cap F = F \cap E$;
Associativity: $(E \cup F) \cup G = E \cup (F \cup G)$ and $(E \cap F) \cap G = E \cap (F \cap G)$
Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ and $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.

De Morgan's law

$$(\cup_{i=1}^{n} E_i)^c = \cap_{i=1}^{n} E_i^c \qquad (\cup_{i=1}^{\infty} E_i)^c = \cap_{i=1}^{\infty} E_i^c$$

$$(\cap_{i=1}^{n} E_i)^c = \cup_{i=1}^{n} E_i^c \qquad (\cap_{i=1}^{\infty} E_i)^c = \cup_{i=1}^{\infty} E_i^c$$

Properties of set theoretical operations

Commutativity: $E \cup F = F \cup E$ and $E \cap F = F \cap E$;
Associativity: $(E \cup F) \cup G = E \cup (F \cup G)$ and $(E \cap F) \cap G = E \cap (F \cap G)$
Distributivity: $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$ and $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$.

De Morgan's law

$$(\cup_{i=1}^{n} E_i)^c = \cap_{i=1}^{n} E_i^c \qquad (\cup_{i=1}^{\infty} E_i)^c = \cap_{i=1}^{\infty} E_i^c$$
$$(\cap_{i=1}^{n} E_i)^c = \cup_{i=1}^{n} E_i^c \qquad (\cap_{i=1}^{\infty} E_i)^c = \cup_{i=1}^{\infty} E_i^c$$

Outline

Consider a random experiment whose sample space is S. A real-valued function P on the space of all events of the experiment is called a probability (measure) if

(1) for all event E, $0 \leq P(E) \leq 1$;

(2) $P(S) = 1$;

(3) for any sequence $E_1, E_2, \ldots$ of mutually disjoint events,

$$P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i).$$

For any event E, $P(E)$ is referred to as the probability of the event E.

Consider a random experiment whose sample space is S. A real-valued function P on the space of all events of the experiment is called a probability (measure) if

(1) for all event E, $0 \leq P(E) \leq 1$;

(2) $P(S) = 1$;

(3) for any sequence $E_1, E_2, \ldots$ of mutually disjoint events,

$$P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i).$$

For any event E, $P(E)$ is referred to as the probability of the event E.

Tossing a fair coin. $P(H) = P(T) = \frac{1}{2}$.

Tossing a coin for which Heads is twice likely as Tails. $P(H) = \frac{2}{3}$, $P(T) = \frac{1}{3}$.

Tossing a fair die. $P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$.

Tossing a fair coin twice: $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$.

Tossing a fair coin. $P(H) = P(T) = \frac{1}{2}$.

Tossing a coin for which Heads is twice likely as Tails. $P(H) = \frac{2}{3}$, $P(T) = \frac{1}{3}$.

Tossing a fair die. $P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$.

Tossing a fair coin twice: $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$.

Tossing a fair coin. $P(H) = P(T) = \frac{1}{2}$.

Tossing a coin for which Heads is twice likely as Tails. $P(H) = \frac{2}{3}$, $P(T) = \frac{1}{3}$.

Tossing a fair die. $P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$.

Tossing a fair coin twice: $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$.

Tossing a fair coin. $P(H) = P(T) = \frac{1}{2}$.

Tossing a coin for which Heads is twice likely as Tails. $P(H) = \frac{2}{3}$, $P(T) = \frac{1}{3}$.

Tossing a fair die. $P(1) = P(2) = \cdots = P(6) = \frac{1}{6}$.

Tossing a fair coin twice: $P(HH) = P(HT) = P(TH) = P(TT) = \frac{1}{4}$.

Tossing a fair die twice. $P((i,j)) = \frac{1}{36}$, $i,j = 1,\ldots,6$.

Tossing a fair coin until an H appears. $P(H) = \frac{1}{2}$, $P(TH) = \frac{1}{4}$, $P(TTH) = \frac{1}{8}$, $P(TTTH) = \frac{1}{16}, \ldots$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A e^{-t}dt$ for any subset A of $\mathbb{R}_+$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A \lambda e^{-\lambda t}dt$ for any subset A of $\mathbb{R}_+$, where $\lambda > 0$ is a constant.

Tossing a fair die twice. $P((i, j)) = \frac{1}{36}$, $i, j = 1, \ldots, 6$.

Tossing a fair coin until an H appears. $P(H) = \frac{1}{2}$, $P(TH) = \frac{1}{4}$, $P(TTH) = \frac{1}{8}$, $P(TTTH) = \frac{1}{16}, \ldots$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A e^{-t} dt$ for any subset A of $\mathbb{R}_+$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A \lambda e^{-\lambda t} dt$ for any subset A of $\mathbb{R}_+$, where $\lambda > 0$ is a constant.

Tossing a fair die twice. $P((i,j)) = \frac{1}{36}$, $i,j = 1,\ldots,6$.

Tossing a fair coin until an H appears. $P(H) = \frac{1}{2}$, $P(TH) = \frac{1}{4}$, $P(TTH) = \frac{1}{8}$, $P(TTTH) = \frac{1}{16},\ldots$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A e^{-t}dt$ for any subset A of $\mathbb{R}_+$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A \lambda e^{-\lambda t}dt$ for any subset A of $\mathbb{R}_+$, where $\lambda > 0$ is a constant.

Tossing a fair die twice. $P((i,j)) = \frac{1}{36}$, $i, j = 1, \ldots, 6$.

Tossing a fair coin until an H appears. $P(H) = \frac{1}{2}$, $P(TH) = \frac{1}{4}$, $P(TTH) = \frac{1}{8}$, $P(TTTH) = \frac{1}{16}, \ldots$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A e^{-t} dt$ for any subset A of $\mathbb{R}_+$.

Measuring the lifetime of a light-bulb. $P(A) = \int_A \lambda e^{-\lambda t} dt$ for any subset A of $\mathbb{R}_+$, where $\lambda > 0$ is a constant.

Outline

Properties of Probability Measures

Suppose that P is a probability measure. Then

(1) $P(\emptyset) = 0$;

(2) if $E_1, \ldots, E_n$ are disjoint, then

$$P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i);$$

(3) if $E \subset F$, the $P(E) \leq P(F)$;

(4) $P(E^c) = 1 - P(E)$;

(5) $P(\cup_{i=1}^n E_i) = 1 - P(\cap_{i=1}^n E_i^c)$;

(6) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Proof

(1) Take $E_1 = E_2 = \cdots = \emptyset$, then

$$P(\emptyset) = P(\cup_{i=1}^{\infty} E_i) = P(\emptyset) + P(\emptyset) + \cdots ,$$

so $P(\emptyset) = 0$.
(2) Take $E_{n+1} = E_{n+2} = \cdots = \emptyset$, then $E_1, E_2, \cdots$ is a sequence of disjoint events, thus by countable additivity,

$$P(\cup_{i=1}^{n} E_i) = P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^{n} P(E_i).$$

(3) $P(F) = P(E) + P(F \setminus E) \geq P(E)$.
(4) $1 = P(E \cup E^c) = P(E) + P(E^c)$.

Proof (cont)

(5) Follows immediately from (4),
(6) Let $I = E \setminus F$, $II = F \setminus E$ and $III = E \cap F$. Then
$P(E \cup F) = P(I \cup II \cup III) = P(I) + P(II) + P(III)$ and
$P(E) = P(I) + P(III)$, $P(F) = P(II) + P(III)$ and $P(E \cap F) = P(III)$.
Thus $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Example

A fair die is tossed 100 times. Find the probability that there is at least one 5.

The complement of "at least one 5" is "there is no 5". So the answer is

$$1 - \left(\frac{5}{6}\right)^{100}.$$

Proof (cont)

(5) Follows immediately from (4),
(6) Let $I = E \setminus F$, $II = F \setminus E$ and $III = E \cap F$. Then
$P(E \cup F) = P(I \cup II \cup III) = P(I) + P(II) + P(III)$ and
$P(E) = P(I) + P(III)$, $P(F) = P(II) + P(III)$ and $P(E \cap F) = P(III)$.
Thus $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Example

A fair die is tossed 100 times. Find the probability that there is at least one 5.

The complement of "at least one 5" is "there is no 5". So the answer is

$$1 - \left(\frac{5}{6}\right)^{100}.$$

Proof (cont)

(5) Follows immediately from (4),
(6) Let $I = E \setminus F$, $II = F \setminus E$ and $III = E \cap F$. Then
$P(E \cup F) = P(I \cup II \cup III) = P(I) + P(II) + P(III)$ and
$P(E) = P(I) + P(III)$, $P(F) = P(II) + P(III)$ and $P(E \cap F) = P(III)$.
Thus $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

Example

A fair die is tossed 100 times. Find the probability that there is at least one 5.

The complement of “at least one 5” is “there is no 5”. So the answer is

$$1 - \left(\frac{5}{6}\right)^{100}.$$

Example

Suppose $P(E) = \frac{1}{2}$, $P(F) = \frac{1}{3}$ and $P(E \cap F) = \frac{1}{4}$. Find (a) $P(E \cup F)$; (b) $P(E \cap F^c)$; (c) $P(E^c \cap F)$; (d) $P(E^c \cap F^c)$;(e) $P(E^c \cup F^c)$.

(a) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$;
(b) $P(E \cap F^c) = P(E) - P(E \cap F)$;
(c) $P(E^c \cap F) = P(F) - P(E \cap F)$;
(d) $P(E^c \cap F^c) = 1 - P(E \cup F)$;
(e) $P(E^c \cup F^c) = 1 - P(E \cap F)$.

Example

Suppose $P(E) = \frac{1}{2}$, $P(F) = \frac{1}{3}$ and $P(E \cap F) = \frac{1}{4}$. Find (a) $P(E \cup F)$; (b) $P(E \cap F^c)$; (c) $P(E^c \cap F)$; (d) $P(E^c \cap F^c)$;(e) $P(E^c \cup F^c)$.

(a) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$;
(b) $P(E \cap F^c) = P(E) - P(E \cap F)$;
(c) $P(E^c \cap F) = P(F) - P(E \cap F)$;
(d) $P(E^c \cap F^c) = 1 - P(E \cup F)$;
(e) $P(E^c \cup F^c) = 1 - P(E \cap F)$.

A generalization of (6) to the case of the union of n events is the following inclusion-exclusion formula, which can be proved by induction.

Inclusion-exclusion formula

If $E_1, E_2, \ldots, E_n$ are events, then

$$P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) - \sum_{i_1<i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1<i_2<i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3})$$
$$+ \cdots + (-1)^{k+1} \sum_{i_1<\cdots<i_k} P(\cap_{j=1}^k E_{i_j})$$
$$+ \cdots + (-1)^{n+1} P(\cap_{i=1}^n E_i).$$

A generalization of (6) to the case of the union of n events is the following inclusion-exclusion formula, which can be proved by induction.

Inclusion-exclusion formula

If $E_1, E_2, \ldots, E_n$ are events, then

$$\begin{aligned} P(\cup_{i=1}^n E_i) = &\sum_{i=1}^n P(E_i) - \sum_{i_1<i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1<i_2<i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) \\ &+ \cdots + (-1)^{k+1} \sum_{i_1<\cdots<i_k} P(\cap_{j=1}^k E_{i_j}) \\ &+ \cdots + (-1)^{n+1} P(\cap_{i=1}^n E_i). \end{aligned}$$