Ch1: Axioms of Probability

1.2 Sample space and events

Experiment (eg. Tossing a die)
Outcome (aka. sample point)
Sample space (S): {all possible outcomes}
Event (E): subset of sample space. Event can be ϕ, S, or any combination of possible outcomes.

- Ex1.1: Tossing a coin once.
  - $S = \{H, T\}$
  - Events can be ϕ, H, T, $\{H, T\}$.
- Ex1.2: Flipping a coin and then (1) tossing a die if Tail or (2) flipping a coin again if Head.
  $S = \{T1, T2, T3, T4, T5, T6, HT, HH\}$
- From the set theory aspect,
  - Sample space ~ universal set (of all possible outcomes in the experiment that satisfies the 3 axioms of probability)
  - Event ~ subset (that satisfies the 3 axioms of probability)

Sample space and events

Ex 1.5: A bus with a capacity of 34 passengers stops at a station some time between 11:00 A.M. and 11:40 A.M. every day.

- sample space S of the experiment, consisting of counting the number of passengers on the bus and measuring the arrival time of the bus
  - $S = \{(n, t): 0 \leq n \leq 34, 11 \leq t \leq 11\frac{2}{3}\}$
- event E that the bus arrives between 11:20 A.M. and 11:40 A.M. with 27 passengers.
  - $E = \{(27, t): 11\frac{1}{3} \leq t \leq 11\frac{2}{3}\}$

Sample space and events

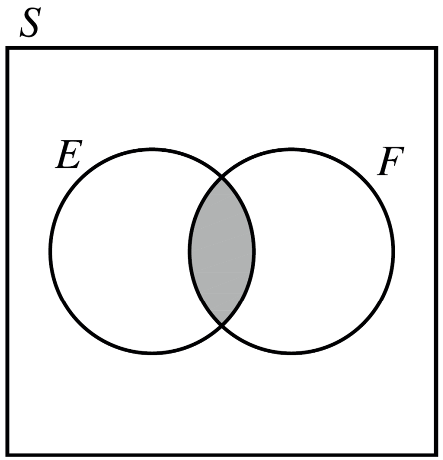
- Event E has occurred in an experiment: If the outcome of an experiment belongs to E.
- Remarks
  - 白馬是馬: Each event belongs to the sample space.
  - 白馬非馬: Not any event is a sample space

Relations between different events

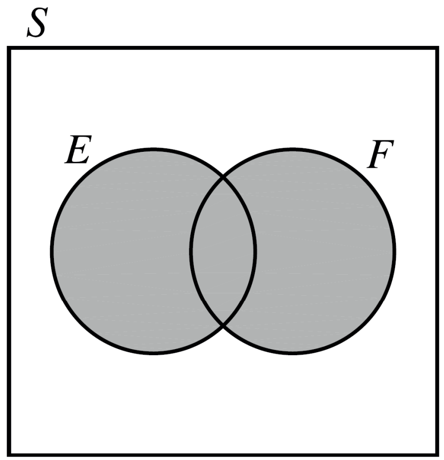
Given the sample space S and two events E and F,

- Subset: $E \subseteq F$ (Whenever E occurs, F also occurs.)
- Equality: if $E \subseteq F$ and $F \subseteq E$, hence $E = F$
- Intersection: $E \cap F$ or EF
- Union: $E \cup F$
- Complement: E^C
- Difference: $E - F$
  - Note:
    - $E^C = S - E$
    - $E - F = E \cap F^C$

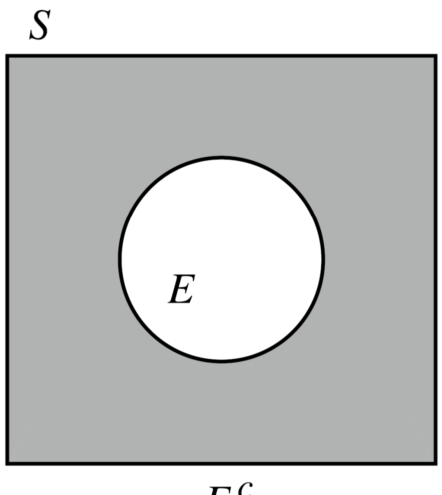
Relations between different events: Venn diagrams



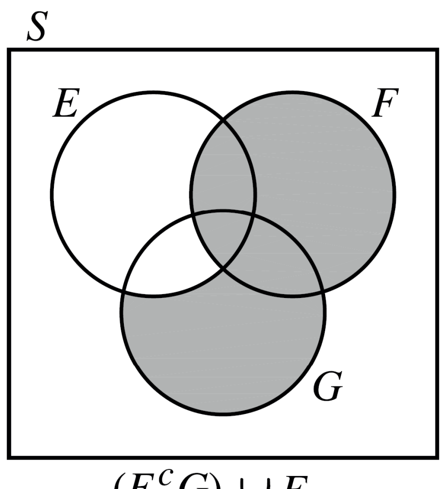
EF



$E \cup F$



E^c



$(E^cG) \cup F$

Relations between different events

- **Certainty**: An event is called **certain** if its occurrence is almost sure (with 100% probability).
  - The sample space is a certain event.
- **Impossibility**: An event is called **impossible** if there is certainty in its nonoccurrence.
  - The empty set $\emptyset$, which is S^c, is an impossible event.

Relations between different events

- **Mutually Exclusiveness:** We say that E and F are **mutually exclusive** if the joint occurrence of E and F is impossible.
  - E and F are mutually exclusive if $EF = \emptyset$.
- A set of events $\{E_1, E_2, \dots\}$ is called **mutually exclusive** if the joint occurrence of any two of them is impossible
  - i.e., $\{E_1, E_2, \dots\}$ is mutually exclusive if $E_iE_j = \emptyset,\ \forall i \neq j$.

Relations between different events

- $\bigcup_{i=1}^{n} E_i$: the union of all the events $E_i, 1 \leq i \leq n$.
  - By $\bigcup_{i=1}^{n} E_i$ we mean the event in which at least one of the events $E_i, 1 \leq i \leq n$, occurs.
- $\bigcap_{i=1}^{n} E_i$: the intersection of all the events $E_i, 1 \leq i \leq n$.
  - $\bigcap_{i=1}^{n} E_i$ occurs when and only when all of the events $E_i, 1 \leq i \leq n$, occur.

Laws and useful relations

- Commutative laws: $E \cup F = F \cup E, EF = FE$
- Associative laws: $E \cup (F \cup G) = (E \cup F) \cup G, E(FG) = (EF)G$
- Distributive laws: $(EF) \cup H = (E \cup H)(F \cup H), (E \cup F)H = EH \cup FH$
- De Morgan's 1st law: $$(E \cup F)^C = E^C F^C, \left(\bigcup_{i=1}^{n} E_i\right)^C = \bigcap_{i=1}^{n} E_i{}^C, \left(\bigcup_{i=1}^{\infty} E_i\right)^C = \bigcap_{i=1}^{\infty} E_i{}^C$$
- De Morgan's 2nd law: $$(EF)^C = E^C \cup F^C, \left(\bigcap_{i=1}^{n} E_i\right)^C = \bigcup_{i=1}^{n} E_i{}^C, \left(\bigcap_{i=1}^{\infty} E_i\right)^C = \bigcup_{i=1}^{\infty} E_i{}^C$$
- $S = F \cup F^c$
- $E = ES = E\left(F \cup F^C\right) = EF \cup EF^C$

Prove De Morgan's first law:

- Show that both $(E \cup F)^C \subseteq E^C F^C$ and $E^C F^C \subseteq (E \cup F)^C$.
- Part I "$(E \cup F)^C \subseteq E^C F^C$"
  - Let x be an outcome in $(E \cup F)^C$.
  - $\because x \notin E \cup F$, $\therefore x \notin E$ and $x \notin F$
  - $x \notin E \Rightarrow x \in E^c$, $x \notin F \Rightarrow x \in F^c$
  - $\because x \in E^c$ and $x \in F^c$, $\therefore x \in E^c \cup F^c$
- Part II "$E^C F^C \subseteq (E \cup F)^C$" is left as hw

1.3 Axioms of probability

Definition (Probability Axioms):

- $\boldsymbol{S}$: the sample space of a random phenomenon
- $\boldsymbol{A}$: an event of S. (Consider each event A.)
- $\boldsymbol{P}$: a real-valued function for each event A,
- If P satisfies the following axioms, then it is called a **probability** and the number $P(A)$ is said to be the **probability of $\boldsymbol{A}$**, for each event A.

1.3 Axioms of probability

- Axiom 1: $P(A) \geq 0$ for each event A
- Axiom 2: $P(S) = 1$
- Axiom 3: If $\{A_1, A_2, \dots\}$ is a sequence of mutually exclusive events (i.e. the joint occurrence of every pair of them is impossible: $A_i A_j = \emptyset$ when $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

1.3 Axioms of probability

- Ex. Tossing a die in which
  - A sequence of events $A_1, A_2, \ldots$ are defined:
    - $A_1 = \{1,2\}$
    - $A_2 = \{3\}$
    - $A_3 = A_4 = \cdots = \emptyset$
  - By Axiom 3
    - $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} A_i$

$$= \frac{2}{6} + \frac{1}{6} + 0 + 0 + \cdots + 0 = \frac{3}{6}$$

Theorem 1.1

- The probability of the empty set ∅ is 0.
  - That is, $P(\emptyset) = 0.$
- Proof:
  - Let $A_1 = S$ and $A_i = \emptyset$ for $i \geq 2$; then $A_1, A_2, A_3, ...$ is a sequence of mutually exclusive events. By Axiom 3,

$$P(S) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

  - Which implies that $\sum_{i=2}^{\infty} P(\emptyset) = 0$. This is only possible only if $P(\emptyset) = 0$ ($\because$ Axiom 1).

Theorem 1.2

- Let $\{A_1, A_2, \dots, A_n\}$ be a mutually exclusive set of events. Then $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.
- Proof:
  - For $i > n$, let $A_i = \emptyset$. Then $A_1, A_2, A_3, \dots$ is a sequence of mutually exclusive events.
  - From Axiom 3 and Theorem 1.1, we get $P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(A_i) = \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) = \sum_{i=1}^{n} P(A_i)$

Theorem 1.4

- For any event A,
  - $P(A) + P(A^c) = 1$
    - Proof:
      - $\because S = A \cup A^c$, $\therefore P(S) = P(A) + P(A^c)$ by Theorem 1.2.
      - $\because P(S) = 1$, $\therefore P(A) + P(A^c) = 1$
- Note
  - For any event A, $0 \leq P(A) \leq 1$
    - Can be proven by $P(A) + P(A^c) = 1$ and Axiom 1 [$\because P(A^c) \geq 0$].

Equally likely

- Let S be the sample space of an experiment. Let A and B be events of S.
  - A and B are **equally likely** if $P(A) = P(B)$.
- Let ω_1 and ω_2 be outcomes of S.
  - ω_1 and ω_2 are **equally likely** if $P(\{\omega_1\}) = P(\{\omega_2\})$.
- If a sample space contains N outcomes and all outcomes are equally likely to occur, then the probability of each outcome (sample point) is $1/N$.
  - Proof:
    - We know
      - $S = \{S_1, S_2, \ldots, S_N\}$
      - $P(\{S_1\}) = P(\{S_2\}) = \ldots = P(\{S_N\})$
    - $\because P(S) = 1$, and outcomes are mutually exclusive sets,
      $\therefore 1 = P(S) = P(\{S_1, S_2, \ldots, S_N\})$
      $= P(\{S_1\}) + P(\{S_2\}) + \ldots + P(\{S_N\}) = N \cdot P(\{S_1\})$
    - $\therefore P(\{S_1\}) = \frac{1}{N}$, Thus $P(\{S_i\}) = \frac{1}{N}$

Theorem 1.3

- If the sample space S of an experiment has N outcomes/points and all outcomes are equally likely to occur, then for any event A of S,

$$P(A) = \frac{N(A)}{N}$$

where $N(A)$ is the # of outcomes/points of A.

Example 1.12

- An elevator with two passengers stops at the second, third, and fourth floors. If it is equally likely that a passenger gets off at any of the three floors, what is the probability that the passengers get off at different floors?
- Ans:
  - $S = \{a_2b_2, a_2b_3, a_2b_4, a_3b_2, a_3b_3, a_3b_4, a_4b_2, a_4b_3, a_4b_4\}$
  - $A = \{a_2b_3, a_2b_4, a_3b_2, a_3b_4, a_4b_2, a_4b_3\}$
  - $\because N = 9$ and $N(A) = 6$, $\therefore \frac{N(A)}{N} = \frac{2}{3}$

1.4 Basic Theorems

- Theorem 1.4: For any event A,
$$P(A^c) = 1 - P(A)$$
- Proof
  - $\because AA^C = \emptyset, \therefore$ A and A^C are mutually exclusive, thus $P(A \cup A^C) = P(A) + P(A^C)$
  - $A \cup A^C = S$ and $P(S) = 1$
    So $1 = P(S) = P(A \cup A^C) = P(A) + P(A^C)$
  - $P(A^C) = 1 - P(A)$

Theorem 1.5

- If $A \subseteq B$, then
  $$P(B-A) = P(BA^c) = P(B) - P(A)$$
- Proof:
  - From Fig 1.2, $A \subseteq B$ implies $B = (B-A) \cup A$
    - where $(B-A)$ and A are mutually exclusive
  - $\because P(B) = P\big((B-A) \cup A\big) = P(B-A) + P(A)$
    $\therefore P(B-A) = P(B) - P(A)$

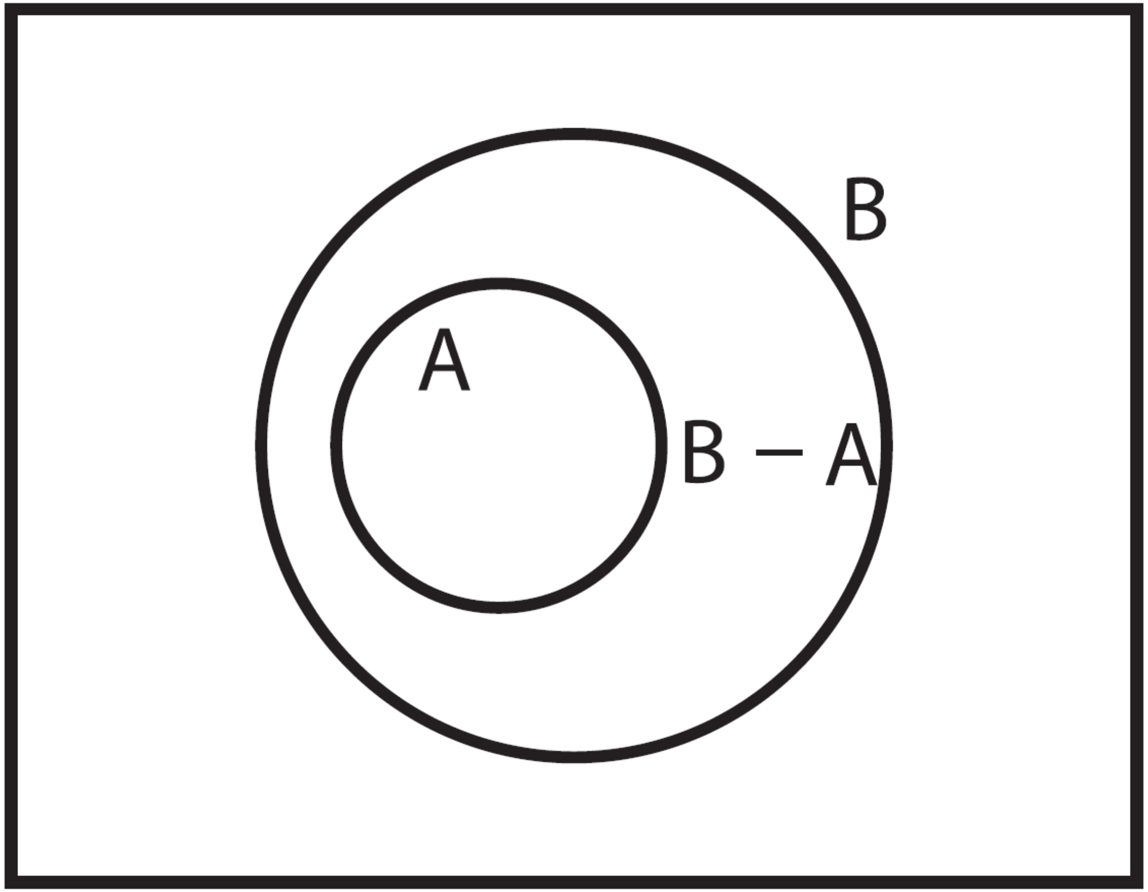


Figure 1.2 $A \subseteq B$ implies that $B = (B - A) \cup A$.

Corollary

- If $A \subseteq B$, then $P(A) \leq P(B)$.
- Proof:
  - By Theorem 1.5, $P(B - A) = P(B) - P(A)$
  - $P(B - A) \geq 0 \implies P(B) - P(A) \geq 0$
  - $\implies P(B) \geq P(A)$

Theorem 1.6

- $P(A \cup B) = P(A) + P(B) - P(AB)$
- Proof:
  1. $A \cup B = A \cup (B - AB)$ ⋯⋯⋯⋯See Fig.1.3
  2. $A(B - AB) = \emptyset$ ⋯⋯mutually exclusive
  3. $P(A \cup B) = P\big(A \cup (B - AB)\big) = P(A) + P(B - AB)$
  4. $AB \subseteq B$, Theorem 1.5 implies that
     $$P(B - AB) = P(B) - P(AB)$$
  5. $P(A \cup B) = P(A) + P(B) - P(AB)$

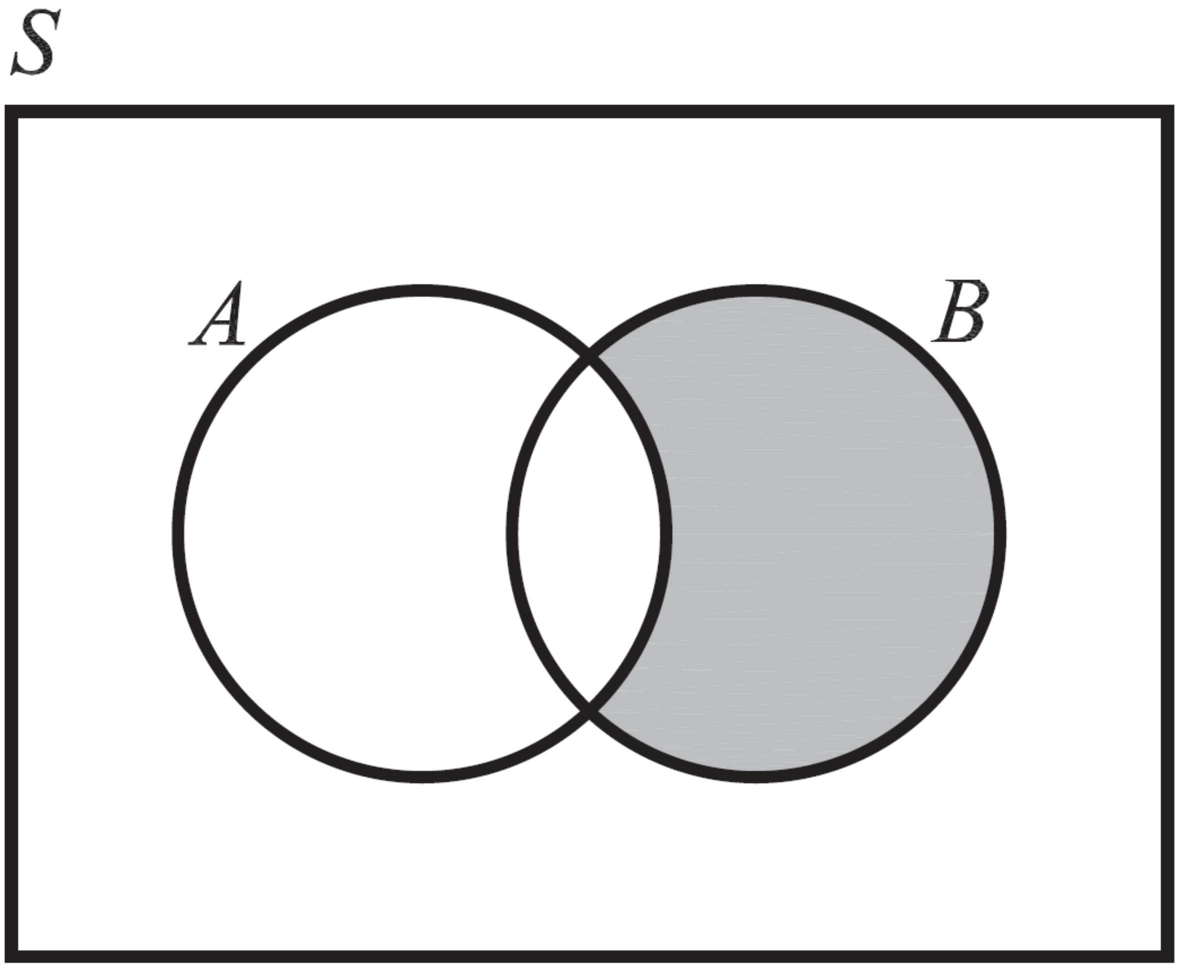


Figure 1.3 The shaded region is $B - AB$. Thus $A \cup B = A \cup (B - AB)$.

Inclusion-Exclusion Principle

$$P(\bigcup_{i=1}^{n} A_i)$$
$$= \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} P(A_i A_j A_k)$$
$$- \cdots + (-1)^{n-1} P(A_1 A_2 A_3 \ldots A_n)$$

- Ex.

$$P(A \cup B \cup C) = [P(A) + P(B) + P(C)] - [P(AB) + P(AC) + P(BC)] + [P(ABC)]$$

Example 1.18

- 25% of the population of a city read newspaper A, 20% read newspaper B, 13% read C, 10% read both A and B, 8% read both A and C, 5% read B and C, and 4% read all three. If a person from this city is selected at random, what is the probability that he or she does not read any of these newspapers?

1. E, F, G : events that the person reads A, B, and C.
2. $P(E \cup F \cup G)$

   $= P(E) + P(F) + P(G) - P(EF) - P(FG) - P(EG) + P(EFG)$

   $= 0.25 + 0.2 + 0.13 - 0.1 - 0.08 - 0.05 + 0.04$

   $= 0.39$
3. $P\{\text{read none of } A, B, C\} = 1 - 0.39 = 0.61$

Theorem 1.7

- $P(A) = P(AB) + P(AB^c)$.
- Proof:

1. $A = AS = A(B \cup B^c) = AB \cup AB^c$
2. AB and AB^c are mutually exclusive
3. $P(A) = P(AB \cup AB^c) = P(AB) + P(AB^c)$

Example 1.19

- In a community, 32% of the population are male smokers; 27% are female smokers. What percentage of the population of this community smoke?
  1. A: event that a randomly selected person from this community smokes.

     B: event that the person is male.
  2. $P(A) = P(AB) + P(AB^c) = 0.32 + 0.27 = 0.59$

1.5 Continuity of probability function

- $f: \mathbb{R} \to \mathbb{R}$ is called continuous at a point $c \in \mathbb{R}$ if $\lim_{x \to c} f(x) = f(c)$
  - $\mathbb{R}$ denotes the set of all real numbers.
- $f: \mathbb{R} \to \mathbb{R}$ is continuous on $\mathbb{R}$ if and only if
$$\lim_{n \to \infty} f(x_n) = f\left(\lim_{n \to \infty} x_n\right)$$
for every convergent sequence $\{x_n\}_{n=1}^{\infty}$ in $\mathbb{R}$.

Increasing

- A sequence $\{E_n, n \geq 1\}$ of events of a sample space is called **increasing** if

  $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots \subseteq E_n \subseteq E_{n+1} \subseteq \cdots$
- For an increasing sequence of events $\{E_n, n \geq 1\}$, by $\lim_{n\to\infty} E_n$ we mean the event that **at least** one $E_i, 1 \leq i < \infty$, occurs. Therefore, $\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$

Decreasing

- A sequence $\{E_n, n \geq 1\}$ of events of a sample space is called **decreasing** if

  $E_1 \supseteq E_2 \supseteq E_3 \supseteq \cdots \supseteq E_n \supseteq E_{n+1} \supseteq \cdots$
- For a decreasing sequence of events $\{E_n, n \geq 1\}$, by $\lim_{n\to\infty} E_n$ we mean the event that **every** E_i occurs.

  Therefore, $\lim_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} E_n$

Theorem 1.8

- **(Continuity of Probability Function)** *For any increasing or decreasing sequence of events,* $\{E_n, n \geq 1\}$*,*

$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n)$$

1. Let $F_1 = E_1, F_2 = E_2 - E_1, \ldots, F_n = E_n - E_{n-1}$
2. $\{F_n, n \geq 1\}$ is a mutually exclusive set of events that satisfies

   (1) $\bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i = E_n, n = 1,2,\ldots$

   (2) $\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i$

Theorem 1.8

3. If $\{E_n, n \geq 1\}$ is increasing

$$P(\lim_{n\to\infty} E_n) = P(\bigcup_{i=1}^{\infty} E_i) = P(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} P(F_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(F_i)$$

$$= \lim_{n\to\infty} P(\bigcup_{i=1}^{n} F_i) = \lim_{n\to\infty} P(\bigcup_{i=1}^{n} E_i) = \lim_{n\to\infty} P(E_n)$$

4. If $\{E_n, n \geq 1\}$ is decreasing

$$P(\lim_{n\to\infty} E_n) = P(\bigcap_{i=1}^{\infty} E_i) = 1 - P[(\bigcap_{i=1}^{\infty} E_i)^c] = 1 - P(\bigcup_{i=1}^{\infty} E_i^c)$$

$$= 1 - P\left(\lim_{n\to\infty} E_n^c\right) = 1 - \lim_{n\to\infty} P(E_n^c)$$

$$= 1 - \lim_{n\to\infty} [1 - P(E_n)] = 1 - 1 + \lim_{n\to\infty} P(E_n) = \lim_{n\to\infty} P(E_n)$$

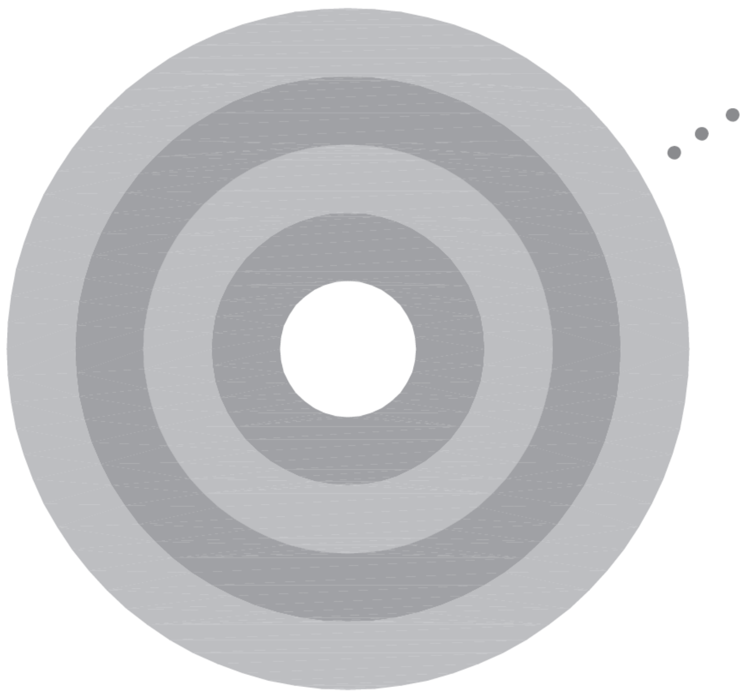


Figure 1.5 The circular disks are the E_i's and the shaded circular annuli are the F_i's, except for F_1, which equals E_1.

Example 1.20

- Suppose that some individuals in a population produce offspring of the same kind.
- If with prob $\exp(-\frac{2n^2+7}{6n^2})$ the entire population completely dies out by the nth generation before producing any offspring.
- What is the prob that such a population survives forever?

Example 1.20 (cont'd)

- Ans: Let En denote the event of extinction of the entire population by the nth generation
- Then $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots \subseteq E_n \subseteq E_{n+1} \subseteq \cdots$;
  because if E_n occurs then E_{n+1} occurs.
  Hence by Theorem 1.8
  p[survives forever] = 1 – p[eventually dies out]

$$= 1 - p\left(\bigcup_{i=1}^{\infty} E_i\right)$$

$$= 1 - \lim_{n\to\infty} P(E_i)$$

$$= 1 - \lim_{n\to\infty} \exp\left(-\frac{2n^2+7}{6n^2}\right) = 1 - e^{-1/3}$$

1.6 Probabilities 0 and 1

- If E and F are events with probabilities 1 and 0, respectively,
  - it is not correct to say that E is the sample space S and F is the empty set $\emptyset$.

Example of zero probability

- Suppose that an experiment consists of selecting a random point from the interval $[0, 1)$.
  - $P\left(\frac{1}{3} \text{ is selected}\right) = 0$
- To show the zero probability:
  - Since every point in $[0, 1)$ has a decimal representation such as 0.xxxxx⋯, the experiment is equivalent to picking an endless decimal from [0, 1) at random.
    - If a decimal terminates, all of its digits from some point on are 0.
  - In such an experiment we want to compute the probability of selecting the point 1/3.
  - In other words, we want to compute the probability of choosing 0.33333⋯ in a random selection of an endless decimal.
    - $P\left(\frac{1}{3} \text{ is selected}\right) = \frac{1}{10} \cdot \frac{1}{10} \cdots = 0$

Remarks

- Examples of zero probability
  - $P\left(\frac{1}{3} \text{ is selected from } [0,1)\right) = 0$
  - $P(\text{falls into any certain value in an interval}) = 0$
- Examples of one probability when $S = [0,1)$
  - $P\{\text{falls into } [0,1) \text{ but not } 0.5\} = P(S - \{0.5\})$
    $= 1 - P(\text{falls into } 0.5) = 1 - 0 = 1$
  - $P(\text{falls into any but not a specific point}) = 1$
  - $P(\text{falls into any but not a few specific points}) = 1$

1.7 Random selection of points from intervals

- What is the occurrence probability of an sub-interval when all random points selected from an interval have a fixed precision?
  - For example, a decimal fraction of 2 digits, i.e. 0.00, 0.01, …, 0.99
    - 100 numbers in total in $[0,1)$
  - $P\{\text{falls in } [0.00, 0.10)\} = 10/100$
    $P\{\text{falls in } [0.00, 0.20)\} = 20/100$
  - Note: The probability is proportional to the length.

Definition

- A point is said to be **randomly selected from an interval** (a, b) if
  - any two subintervals of (a, b) that have the same length are equally likely to include the point.
- The probability associated with the event that the subinterval (α, β) contains the point is defined to be $\frac{\beta-\alpha}{b-a}$.

Open vs. closed sub-intervals

- $P\{[\alpha,\beta)\}$ vs. $P\{(\alpha,\beta)\}$

$$P\{[\alpha,\beta)\} = P\{(\alpha,\beta)\} + P\{\alpha\}$$
$$= P\{(\alpha,\beta)\}$$

- Similarly,
  - $P\{(\alpha,\beta)\} = P\{[\alpha,\beta)\} = P\{(\alpha,\beta]\} = P\{[\alpha,\beta]\}$