METU MATH DEPT. MATH 401 Video 1

Probability Theory – Fall 2021

Note Title 11.10.2021

Textbook: Elements of the Probability Theory (2nd Ed.)

by H. Koralloğlu, A.B. Hayfavi and Y.Y. Okur

CHAPTER 1:

§1.1. **Introduction**: There are two types of experiments:

- First kind with deterministic outcome; for example water boils at 100°C.

- Second kind with random outcome; for example tossing a coin, rolling a dice and drawing a card from a pack.

Probability theory deals with the experiments of the second kind, i.e., experiments with random outcomes. A random event is a possible outcome of an "experiment" whose occurence is uncertain. Here an "experiment" will mean that not only natural phenomena but also the ones provoked by human beings. Some examples of random events are as follow:

1) the weather in a future time
2) the outcome of elections
3) the change of wining at black jack (21 or "Bazbuk")
4) population growth
5) the evolution of stock market
6) drawing a particular card from a pack.

Randomness: Given an event A, let $n(A)$ denote the number of occurence of A in n trials. Then $0 \leq n(A) \leq n$. The ratio $n(A)/n$ is called the relative frequency and $0 \leq n(A)/n \leq 1$. If the limit

$$\lim_{n \to \infty} \frac{n(A)}{n} \quad \text{exists we denote it by } \mathbb{P}(A)$$

and we say that the experiment has statistical stability. Probability theory deals with mathematical models of experiments with random outcomes having statistical stability.

§1.2. Basic Concepts of Probability Theory:

Sample Space: A sample space is a non-empty set of points, denoted by Ω and its elements by ω.

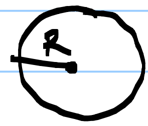
Examples: 1) Tossing a coin, $\Omega = \{h, t\}$

2) Rolling a die $\Omega = \{1, 2, 3, 4, 5, 6\}$.

3) Rolling a pair of dice

$$\Omega = \{(1,1), (1,2), \dots, (1,6), (2,1), \dots, (2,6), \dots, (6,1), \dots, (6,6)\}.$$

4) Shooting a round target with radius R



$$\Omega = \{(x,y) \mid x^2+y^2 \leq R^2\}.$$

Algebra of Events: An event is a possible outcome of an experiment. Unless stated otherwise, an event and its realization will have the same meaning. The occurrence of an event may depend on that of other events.

Example: Suppose two dice are thrown. The event "the sum of the numbers shown up by each die is equal to 10". Let X and Y denote the number shown by the first and the second die, respectively.

Video 2

Let E denote the event $\{X+Y=10\}$. Then E occurs if and only if either $E_1=\{\{X=4\}$ and $\{Y=6\}\}$ or $E_2=\{\{X=5\}$ and $\{Y=5\}\}$ or $E_3=\{\{X=6\}$ and $\{Y=4\}\}$ occurs. Then we write this as

$E=(E_1$ or E_2 or $E_3)$.

Notation: Let A and B be events. Then $A\cup B$ will denote the event $(A$ or $B)$ and $A\cap B$ will denote the event $(A$ and $B)$. Finally, A^c will mean the non-occurrence of A.

If Ω is a sample space, i.e., an abstract set of points, which are considered as a possible outcome of an experiment, then any subset A of Ω is called an event. The nonempty collection of events will be denoted as $\mathcal{A}$.

Convention: If Ω is a finite set, then $\mathcal{A}$ will denote the power set of Ω, i.e., the set of all subsets of Ω.

$|\Omega|<+\infty$, $\mathcal{A}=\mathcal{P}(\Omega)$.

On the other hand, if Ω is infinite it is not reasonable to take $\mathcal{A}$ as the power set of Ω. However, will expect that the collection $\mathcal{A}$ of events satisfy the following conditions:

i) If $A\in\mathcal{A}$, then so is A^c.
ii) If A and B are in $\mathcal{A}$ then so are $A\cap B$ and $A\cup B$.

Note that by induction we see that if $A_1,\dots,A_n$ are sets in $\mathcal{A}$ then so are $\bigcup_{i=1}^{n}A_i$ and $\bigcap_{i=1}^{n}A_i$.

Note that if $A \in \mathcal{A}$ then $A \cap A^c = \emptyset \in \mathcal{A}$ and $A \cup A^c = \Omega \in \mathcal{A}$.

<u>Definition</u> A Boolean algebra (a field of subsets of Ω or algebra of events) is a nonempty set $\mathcal{A}$ on which two binary operations $\cup, \cap$ and one unary operation ^{c} are defined so that

$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$ and $A \cap B \in \mathcal{A}$

$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ and satisfying the following axioms:

a_1) $A \cup B = B \cup A$, $A \cap B = B \cap A$

a_2) $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$

a_3) $(A \cap B) \cup B = B$, $(A \cup B) \cap B = B$.

a_4) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

a_5) $(A \cap A^c) \cup B = B$, $(A \cup A^c) \cap B = B$.

<u>Example:</u> Let $\Omega = \{1,2,3,4\}$.

$\mathcal{A}_1 = \{\emptyset, \{1,2\}, \{3,4\}\}$, $\mathcal{A}_2 = \{\emptyset, \Omega, \{1\}, \{2,3,4\}, \{1,2\}, \{3,4\}\}$,

$\mathcal{A}_3 = \{\emptyset, \Omega, \{1\}, \{2\}, \{1,2\}, \{3,4\}, \{2,3,4\}, \{1,3,4\}\}$

$\mathcal{A}_4 = \mathcal{P}(\Omega)$

Then $\mathcal{A}_3$ and $\mathcal{A}_4$ are Boolean algebras but $\mathcal{A}_1$ and $\mathcal{A}_2$ are not.

<u>Remarks</u> 1) Using mathematical induction one can prove that for any given events $A_1, \dots, A_n \in \mathcal{A}$ the axiom (a_2) we may drop the parenthesis and $A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^{n} A_i$ and $A_1 \cap \dots \cap A_n = \bigcap_{i=1}^{n} A_i$.

2) (Idempotent laws) For all $A \in \mathcal{A}$, $A \cup A = A$ and $A \cap A = A$.

3) For $A, B \in \mathcal{A}$, note that $A \cap B = B \Leftrightarrow A = A \cup B \Leftrightarrow B \subseteq A$.

4) $A \subseteq A$ (reflexivity)

5) $A \subseteq B$, $B \subseteq A \Rightarrow A = B$ (anti symmetry)

6) $A \subseteq B$, $B \subseteq C \Rightarrow A \subseteq C$ (transitivity).

The element $\phi \in \mathcal{A}$ is called the null element and Ω is called the unit element.

Note that $A \cap A^c = \phi$ and $A \cup A^c = \Omega$, for any $A \in \mathcal{A}$.

7) Note that for any given $A \in \mathcal{A}$, $A^c \doteq \Omega \setminus A$ is the unique element X satisfying $X \cap A = \phi$ and $X \cup A = \Omega$.

8) The element $A \setminus B \doteq A \cap B^c$ is called the difference of B from A.

9) $A \triangle B \doteq (A \setminus B) \cup (B \setminus A)$ is called the symmetric difference of A and B.

Definition: An element $A \in \mathcal{A}$ is called an atom of $\mathcal{A}$ if $A \neq \phi$ and ($B \subseteq A$, $B \in \mathcal{A}$, then $B = \phi$ or $B = A$).

Note that if A and B are distinct atoms of $\mathcal{A}$

Video 3

then $A \cap B = \emptyset$.

Example: $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{3,4\}, \{2,3,4\}, \{1,3,4\}, \{1,2,3,4\}\} \subseteq \mathcal{P}(\Omega)$, $\Omega = \{1,2,3,4\}$.

$\{1\}$, $\{2\}$ and $\{3,4\}$ are atoms of $\mathcal{A}$.

Similarly, if $\mathcal{B} = \{\emptyset, (0,1), (0, 2/3), [2/3, 1)\}$ ($\Omega = (0,1)$) then $(0, 2/3)$ and $[2/3, 1)$ are atoms of $\mathcal{B}$.

Example: 30, $\mathcal{B} = \{1, 2, 3, 5, 6, 10, 15, 30\}$

$x, y \in \mathcal{B}$, $x \cup y = \text{l.c.m}(x,y)$, $x \cap y = \text{g.c.d.}(x,y)$.
$x^c = \frac{30}{x}$.

$3 \cup 5 = 15$, $6 \cap 10 = 2$, $10^c = \frac{30}{10} = 3$.

Proposition: Let $\mathcal{A}$ be a finite Boolean algebra. Then every element A of $\mathcal{A}$ is a union of the atoms that are contained in A.

Proof: Ω finite set, $\mathcal{A} \subseteq \mathcal{P}(\Omega)$.

Let $K = \{B_1, \ldots, B_n\}$ be the set of atoms of $\mathcal{A}$. Then $\Omega = B_1 \cup \cdots \cup B_n$. Now take any $A \in \mathcal{A}$. If $A \cap B_i = \emptyset$ for all $i = 1, \ldots, n$, then $A \subseteq B_1^c \cap \cdots \cap B_n^c = (B_1 \cup \cdots \cup B_n)^c = \Omega^c = \emptyset$. Hence $A = \emptyset$.

Hence, any nonempty $A \in \mathcal{A}$ must intersect some B_i. If $A \cap B_i \neq \emptyset$ then $B_i \subseteq A$ because otherwise $B_i = (A \cap B_i) \cup (B_i \cap A^c)$, which
$\neq \emptyset$ $\neq \emptyset$

implies that B_i is not an atom.

Let $A' = \bigcup_{B_i \subseteq A} B_i$. Clearly, $A' \subseteq A$. If $A \setminus A' \neq \emptyset$ then $(A \setminus A') \cap B_j \neq \emptyset$ for some $j \in \{1, \dots, n\}$. This implies $B_j \subseteq A \setminus A'$. Now $B_j \subseteq A$ and $B_j \cap A' = \emptyset$, a contradiction to the definition of A'. Hence $A = A' = \bigcup_{B_i \subseteq A} B_i$.

<u>Corollary</u> Assume $\mathcal{A}$ is a finite Boolean algebra and $K = \{B_1, \dots, B_n\}$ is the set of atoms of $\mathcal{A}$. Then $\mathcal{A}$ has 2^n elements.

<u>Definition</u>: Let $\mathcal{A}$ be a Boolean algebra on Ω. A sub-Boolean algebra of $\mathcal{A}$ is a Boolean algebra $\mathcal{B}$ so that $\Omega \in \mathcal{B}$ and $\mathcal{B} \subseteq \mathcal{A}$.

Given a subset $\mathcal{C}$ of $\mathcal{A}$, the smallest sub-Boolean algebra containing $\mathcal{C}$ is said to be generated by $\mathcal{C}$.

<u>Exercise</u>: Show that the intersection of any collection $\{\mathcal{B}_\lambda\}_{\lambda \in \Lambda}$ of sub-Boolean algebras of $\mathcal{A}$ is a sub-Boolean algebra of $\mathcal{A}$.

<u>Proposition</u>: If $B_1, \dots, B_n$ are distinct elements of a Boolean algebra $\mathcal{A}$ then the atoms of the sub-Boolean algebra $\mathcal{B}$ generated by $B_1, \dots, B_n$ are of the form $\overline{B_{i_1} \cap \dots \cap B_{i_k}}$, for some $1 \leq i_1 < i_2 < \dots < i_k \leq n$ and therefore $\mathcal{B}$ has at most 2^{2^n} elements.

Video 4

Example: Let A and B be two distinct elements of a Boolean algebra. Then the sub-Boolean algebra generated by A and B is the algebra generated by $A\cap B$, $A\cap B^c$, $A^c\cap B$ and $A^c\cap B^c$, if they do not reduce to ϕ.

In this case, $\mathcal{B}=\{\phi, \Omega, A\cap B, A\cap B^c, A^c\cap B, A^c\cap B^c, A, B, A\cup B, A\cup B^c, B\cup A^c, \ldots\}$.

Remark: The Boolean algebra described above is indeed standard:

$\mathcal{B}=\{1,2,3,5,6,10,15,30\}$, $x,y\in\mathcal{B}$

$x\cup y = \text{l.c.m.}\{x,y\}$, $x\cap y=\text{g.c.d.}\{x,y\}$

$x^c=\dfrac{30}{x}$

This algebra is isomorphic to the following standard algebra:

$\Omega'=\{2,3,5\}$ $\quad \mathcal{A}=\mathcal{P}(\Omega')$
$=\{\phi, \{2\}, \{3\}, \{5\}, \ldots, \{2,3,5\}\}$

If $A,B\in\mathcal{A}$ let $A\cup B$, $A\cap B$ the usual set union and intersection.

Isomorphism: $\mathcal{A}\longleftrightarrow\mathcal{B}$

$\phi \longleftrightarrow 1$

$\{2\}\longleftrightarrow 2$

$\{2,3\}\longleftrightarrow 2\cdot 3=6$

$\{3,5\}\longleftrightarrow 3\cdot 5=15$

$A = \{2\}$, $B = \{3\}$

$A \cup B = \{2, 3\} \longleftrightarrow 6 = 2 \cdot 3 = \text{l.c.m.}\{2, 3\}$

$A = \{2, 3\}$, $B = \{3, 5\}$, $A \cap B = \{3\}$

$3 = \text{g.c.d.}\{6, 15\}$

Sigma Algebra of Events.

Definition: Let Ω be a sample set. A nonempty class $\mathcal{A}$ of subsets of Ω is called a σ-algebra if

σ_1) $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$

σ_2) $A \in \mathcal{A}$ then $A^c = \Omega \setminus A \in \mathcal{A}$

σ_3) If $\{A_i\}_{i \in I}$ is a countable class of elements of $\mathcal{A}$, then $\bigcup_{i \in I} A_i \in \mathcal{A}$ (or equivalently, $\bigcap_{i \in I} A_i \in \mathcal{A}$).

The elements of $\mathcal{A}$ are called events and, in Measure Theory, they are called measurable sets.

Definition: Let Ω be a sample space and let $\mathcal{A}$ be a σ-algebra of subsets of Ω. Then the pair $(\Omega, \mathcal{A})$ is called a measurable space.

The element Ω is called the sure event and $\emptyset$ the impossible event.

Examples 1) $\{\emptyset, \Omega\}$ is the simplest σ-algebra

2) $\{\emptyset, \Omega, A, A^c = \Omega \setminus A\}$ is the σ-algebra generated by the event A.

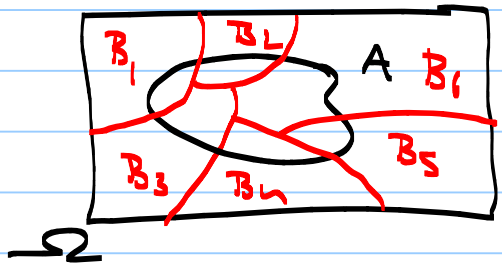
3) $\{\emptyset, \Omega, A, B, A\cap B, A^c, B^c, A^c \cap B^c, A\cup B^c, A\cap B^c, B\cap A^c,$
$C=(A^c\cap B)\cup(A\cap B^c), A\cup B, A^c\cup B^c, A^c\cup B, C^c\}$ is the σ-algebra generated by the events A and B.

4) The family $\mathcal{P}(\Omega)$ of all subsets of Ω, is the biggest σ-algebra with the sample space Ω.

<u>Definition:</u> Let $\{B_i \mid i\in I\}$ be a countable family of (non-empty) events in the σ-algebra $\mathcal{A}$. The $\{B_i \mid i \in I\}$ is called a measurable partition of Ω if B_i's are pairwise disjoint and $\bigcup_{i\in I} B_i = \Omega$.

<u>Remark:</u> If $\{B_i \mid i\in I\}$ is a measurable partition of Ω then for any event $A\in\mathcal{A}$ we have

$$A = \bigcup_{i\in I} (A\cap B_i)$$



<u>Example:</u> In the experiment of tossing a pair of dice $D_i = \{(j,k) \mid 1\le j,k\le 6,\ j+k=i\}$ $i=2,3,\dots,12$ is a measurable partition of Ω.

$\Omega = \{(i,j) \mid 1\le i,j\le 6\}$.

$\Omega = \bigcup_{i=2}^{12} D_i$, $\quad D_2=\{(1,1)\}$, $D_{10}=\{(4,6),(5,5),(6,4)\}$

Video 5

<u>Exercise</u>: If $\{\mathcal{A}_i \mid i \in I\}$ is a family of σ-algebras of subsets of a sample space Ω, then show that $\bigcap_{i \in I} \mathcal{A}_i$ is also a σ-algebra.

<u>Definition</u>: Let $\mathcal{C}$ be a subset of $\mathcal{P}(\Omega)$ of a sample space Ω. The smallest σ-algebra $\sigma(\mathcal{C})$ containing $\mathcal{C}$, is the intersection all σ-algebras containing $\mathcal{C}$. This is called the σ-algebra generated by $\mathcal{C}$.

<u>Definition</u>: If $\mathcal{A}$ and $\mathcal{C}$ are σ-algebras and $\mathcal{C} \subseteq \mathcal{A}$ then $\mathcal{C}$ is called a sub-σ-algebra of $\mathcal{A}$.

<u>Definition</u>: Let $\mathcal{G}$ (respectively $\mathcal{F}$) be the family of all open (respectively, closed) subsets of $\mathbb{R}^n$. Then $\sigma(\mathcal{G}) = \sigma(\mathcal{F})$ is called the Borel σ-algebra of $\mathbb{R}^n$. We'll denote it by $\mathcal{B}^n$, for $n \geq 2$ and $\mathcal{B}$ for $n = 1$. The elements of $\mathcal{B}^n$ are called Borel sets of $\mathbb{R}^n$.

<u>Proposition</u>: $\mathcal{B}^n$ is generated by the countable family of products of intervals

$$S = \{(-\infty, x_1] \times \cdots \times (-\infty, x_n] \mid (x_1, \dots, x_n) \in \mathbb{Q}^n\}.$$

<u>Proof</u>: 1) Note that products of the form

$$\mathbb{R} \times \mathbb{R} \times \cdots \times (-\infty, x] \times \mathbb{R} \times \cdots \times \mathbb{R} \subseteq \mathbb{R}^n$$ belongs to $\sigma(S)$ for any $x \in \mathbb{Q}$.

$$\mathbb{R} \times \cdots \times (-\infty, x] \times \mathbb{R} \cdots \times \mathbb{R} = \bigcup_{n=1}^{\infty} \underbrace{(-\infty, n] \times \cdots \times (-\infty, x] \times \cdots \times (-\infty, n]}_{\in S}$$

2) $\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,x]\times\mathbb{R}\cdots\times\mathbb{R}\in\sigma(S)$, for any $x\in\mathbb{R}$.
If $x\in\mathbb{R}$, let (r_n) be a rational sequence with $r_n\leq r_{n+1}$ and $\lim r_n = x$. Then

$$\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,x]\times\mathbb{R}\times\cdots\times\mathbb{R} = \bigcap_{n=1}^{\infty}\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,r_n]\times\mathbb{R}\times\cdots\times\mathbb{R}.$$

3) $(\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,x]\times\cdots\times\mathbb{R})^c = \mathbb{R}\times\cdots\times\mathbb{R}\times(x,\infty)\times\mathbb{R}\times\cdots\times\mathbb{R}$
for any $x\in\mathbb{R}$.
Also,

$$\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,y)\times\cdots\times\mathbb{R} = \bigcup_{n=1}^{\infty}\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,y-1/n]\times\cdots\times\mathbb{R}$$

so that both belong to the Borel algebra $\mathcal{B}^n$.

Finally, $(\mathbb{R}\times\cdots\times\mathbb{R}\times(x,\infty)\times\cdots\times\mathbb{R})$
$\cap(\mathbb{R}\times\cdots\times\mathbb{R}\times(-\infty,y)\times\mathbb{R}\times\cdots\times\mathbb{R})$
$=\mathbb{R}\times\cdots\times\mathbb{R}\times(x,y)\times\mathbb{R}\times\cdots\times\mathbb{R}\in\mathcal{B}^n$,

for any $x,y\in\mathbb{R}$.

4) $(x_1,y_1)\times\cdots\times(x_n,y_n) = [(x_1,y_1)\times\mathbb{R}\times\cdots\times\mathbb{R}]\cap$
$[\mathbb{R}\times(x_2,y_2)\times\cdots\times\mathbb{R}]\cap$
$\vdots$
$[\mathbb{R}\times\cdots\times\mathbb{R}\times(x_n,y_n)]$

So that $(x_1,y_1)\times\cdots\times(x_n,y_n)\in\mathcal{B}^n$, for all $x_i,y_i\in\mathbb{R}$.

5) Let $U\subseteq\mathbb{R}^n$ be any open subset. Take any point $p=(z_1,\dots,z_n)\in U$. Choose an element $B_p=(x_1,y_1)\times\cdots\times(x_n,y_n)\in\mathcal{B}^n$ so that $x_i,y_i\in\mathbb{Q}$,
$z_i\in(x_i,y_i)$, $i=1,\dots,n$, and $(x_1,y_1)\times\cdots\times(x_n,y_n)\subseteq U$.

Then $U=\bigcup_{p\in U}\{p\}\subseteq\bigcup_{p\in U}B_p\subseteq U$ and thus

Video 6

$\mathcal{U} = \bigcup_{p \in \mathcal{U}} B_p$. This is a countable union where

$B_p = (x_1, y_1) \times \cdots \times (x_n, y_n)$, $x_i, y_i \in \mathbb{Q}$.

This finishes the proof. ■

__Definition:__ Let $\mathcal{A}$ be a σ-algebra and $\{A_i, i \in I\}$ be a family of events in $\mathcal{A}$. Then we define $\sup_{i \in I} A_i$ and $\inf_{i \in I} A_i$ as follows:

$$\sup_{i \in I} A_i = \bigcup_{i \in I} A_i \quad \text{and} \quad \inf_{i \in I} A_i = \bigcap_{i \in I} A_i.$$

Similarly, we define for a sequence $(A_n)_{n \in \mathbb{N}}$ of events in $\mathcal{A}$

$$\limsup_n A_n \doteq \bigcap_m \bigcup_{n \geq m} A_n \quad \text{and} \quad \liminf_n A_n = \bigcup_m \bigcap_{n \geq m} A_n.$$

Note that $\limsup A_n$ and $\liminf A_n$ belong to $\mathcal{A}$.

__Remark:__ Let $B_m \doteq \bigcup_{n \geq m} A_n$, then $B_{m+1} \subseteq B_m$ and

$\limsup A_n = \lim_m B_m = \bigcap_{m=1}^{\infty} B_m$. So if $x \in \limsup_n A_n$

then $x \in B_m$ for all m. Hence, $x \in A_n$ for infinitely many $n \in \mathbb{N}$. In other words, the event $\limsup_n A_n$ occurs if and only if infinitely many A_n occurs.

Similarly, if $\underline{\underline{C_m}} \doteq \bigcap_{n \geq m} A_n$, then $C_m \subseteq C_{m+1}$

and $\liminf_n A_n = \lim_m C_m = \bigcup_{m=1}^{\infty} C_m$. Hence, $x \in \liminf A_n$

if and only if $x \in C_m$, for some $m \in \mathbb{N}$.

Hence, $x \in \liminf A_n$ if and only if $x \in A_n$ for all $n \geq m_0$ for some m_0. In other words, the event $\liminf A_n$ occurs if and only if all but finitely many A_n's occur.

Will use the following notation

$\limsup_n A_n = \{A_n, \text{ i.o.}\}$

and

$\liminf_n A_n = \{A_n, \text{ ult.}\}$, where

i.o. stands for "infinitely many A_n occurs" and ult. stand for "A_n occurs ultimately".

Note that if $\{A_n \mid n \in \mathbb{N}\}$ is a non-decreasing ($A_n \subseteq A_{n+1}$ $\forall n$) sequence then we define

$\lim A_n = \bigcup_n A_n$ and if $\{A_n \mid n \in \mathbb{N}\}$ is a non-increasing sequence ($A_{n+1} \subseteq A_n$, $\forall n$) then we define $\lim A_n = \bigcap_n A_n$.

A non-decreasing or non-increasing sequence is called monotone. Clearly, for a monotone sequence $\{A_n \mid n \in \mathbb{N}\}$ we have

$\limsup A_n = \liminf A_n = \lim A_n$.

<u>Proof</u> Assume $\{A_n \mid n \in \mathbb{N}\}$ is non-decreasing. So $A_n \subseteq A_{n+1}$, for all n.

$B_m = \bigcup_{n \geq m} A_n$. Then $B_m \subseteq B_{m+1}$, for all m

$\limsup A_n = \lim B_m = \bigcup_{m=1}^{\infty} B_m = \bigcup_{n=1}^{\infty} A_n = \lim A_n$.

Similarly let $C_m = \bigcap_{n \geq m} A_n = A_m$ and hence,

$\lim\inf A_n = \lim C_m = \lim x_n$.

<u>Definition</u>, A class $\mathcal{C}$ of subsets of Ω is called a monotone class if for any monotone sequence $\{A_n \mid n \in \mathbb{N}\} \subseteq \mathcal{C}$, $\lim\limits_n A_n \in \mathcal{C}$.

<u>Proposition</u>: A Boolean algebra $\mathcal{A}$ is a σ-algebra if and only if it is a monotone class.

<u>Proof</u>: By definition any σ-algebra is a Boolean algebra.

Now assume that $\mathcal{A}$ is a Boolean algebra and it is a monotone class.

<u>must show</u>: $\mathcal{A}$ is a σ-algebra.

It is enough to show the following: let $\{A_n \mid n \in \mathbb{N}\}$ be a countable family in $\mathcal{A}$. We must show that $\bigcup A_n \in \mathcal{A}$.

Let $B_m = \bigcup\limits_{n \leq m} A_n = A_1 \cup A_2 \cup \cdots \cup A_m$. Clearly each $B_m \in \mathcal{A}$ since $\mathcal{A}$ is a Boolean algebra. Also note that $B_m \subseteq B_{m+1}$ for all m, so that $\{B_m \mid m \in \mathbb{N}\}$ is a monotone sequence. Finally, since $\mathcal{A}$ is a monotone class

$$\bigcup_{m=1}^{\infty} B_m = \lim B_m \in \mathcal{A}.$$ However,

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{m=1}^{\infty} B_m = \lim B_m \in \mathcal{A}$$ and the proof finishes. ∎

Let $\mathcal{C}$ be any class of subsets of Ω. Then the σ-algebra generated by $\mathcal{C}$, $\sigma(\mathcal{C})$, is a monotone class containing $\mathcal{C}$. The smallest monotone class containing $\mathcal{C}$ will be denoted as $\mathcal{M}(\mathcal{C})$.

Video 7

Note that $\mathcal{M}(\mathcal{C})$ is the intersection of all monotone classes containing $\mathcal{C}$.

<u>Proposition:</u> (Monotone Class Theorem for Sets) Let $\mathcal{M}$ be a monotone class that contains a Boolean algebra $\mathcal{C}$. Then $\sigma(\mathcal{C}) \subseteq \mathcal{M}$.
Equivalently, for a Boolean algebra $\mathcal{C}$ we have $\sigma(\mathcal{C}) = \mathcal{M}(\mathcal{C})$.

<u>Proof:</u> Let $\mathcal{C}$ be a Boolean algebra. Then by the previous proposition $\sigma(\mathcal{C})$ is a monotone class containing $\mathcal{C}$. Hence, $\mathcal{M}(\mathcal{C}) \subseteq \sigma(\mathcal{C})$ since $\mathcal{M}(\mathcal{C})$ is the smallest monotone class containing $\mathcal{C}$ and $\sigma(\mathcal{C})$ is a monotone class containing $\mathcal{C}$.

<u>Claim</u> $\mathcal{M}(\mathcal{C})$ is a Boolean Algebra.
Note that since $\mathcal{M}(\mathcal{C})$ is a monotone class by the previous proposition, the claim implies that $\mathcal{M}(\mathcal{C})$ is a σ-algebra. Thus we get $\sigma(\mathcal{C}) \subseteq \mathcal{M}(\mathcal{C})$. In particular, $\sigma(\mathcal{M}) \subseteq \mathcal{M}$.

<u>Proof of the claim:</u> It is enough to prove that

i) $\mathcal{M}(\mathcal{C})$ is closed under taking complements, and
ii) $\mathcal{M}(\mathcal{C})$ is closed under finite intersections (or unions).

<u>Proof of i)</u> Let $\mathcal{M}' = \{B \mid B \text{ and } B^c \text{ are in } \mathcal{M}(\mathcal{C})\}$.
Since $\mathcal{C}$ is a Boolean algebra $\mathcal{M}'$ contains $\mathcal{C}$. Let $B_1 \subset B_2 \subset \cdots \subset B_n \subset \cdots$ be a monotone sequence in $\mathcal{M}'$. Then $B_1^c \supset B_2^c \supset \cdots \supset B_n^c \supset \cdots$ is also $\mathcal{M}'$.
Clearly, $\bigcup_n B_n = \lim B_n$ and $\bigcap_n B_n^c = \lim B_n^c$

are both in $\mathcal{M}'$, since $\mathcal{M}(\mathcal{C})$ is a monotone class. Moreover, $(\bigcup_n B_n)^c = \bigcap_n B_n^c$ and thus both lie in $\mathcal{M}'$. In particular, $\mathcal{M}'$ is a monotone class. Since $\mathcal{M}'$ contains $\mathcal{C}$, we have $\mathcal{M}(\mathcal{C}) \subseteq \mathcal{M}'$ so that $\mathcal{M}' = \mathcal{M}(\mathcal{C})$.

In particular, $\mathcal{M}(\mathcal{C})$ is closed under complement.

<u>Proof of ii)</u> Let $A \in \mathcal{C}$ and set

$$\mathcal{M}_A = \{B \in \mathcal{M}(\mathcal{C}) \mid B \cap A \in \mathcal{M}(\mathcal{C})\}.$$

Note that for any $C \in \mathcal{C}$, $C \cap A \in \mathcal{C} \subseteq \mathcal{M}(\mathcal{C})$ and $C \in \mathcal{M}_A$. So $\mathcal{C} \subseteq \mathcal{M}_A$.

Now let $B_1 \subset B_2 \subset B_3 \subset \cdots$ be a sequence in $\mathcal{M}_A$. So $B_1 \cap A \subset B_2 \cap A \subset \cdots \subset B_n \cap A \subseteq \cdots$ is a sequence in $\mathcal{M}(\mathcal{C})$. Since $\mathcal{M}(\mathcal{C})$ is a monotone class $(\bigcup_n B_n) \cap A = \bigcup_n (B_n \cap A) \in \mathcal{M}(\mathcal{C})$. So, $\bigcup_n B_n \in \mathcal{M}_A$.

Similar argument implies that if $B_1 \supset B_2 \supset \cdots \supset B_n$ is a sequence in $\mathcal{M}_A$ then $\lim B_n = \bigcap_n B_n \in \mathcal{M}_A$.

Hence, $\mathcal{M}_A$ is a monotone class.

However, $\mathcal{C} \subseteq \mathcal{M}_A$ and thus $\mathcal{M}(\mathcal{C}) \subseteq \mathcal{M}_A$.

Hence, $\mathcal{M}(\mathcal{C}) = \mathcal{M}_A$.

Now let $A \in \mathcal{M}(\mathcal{C})$ and set

$$\mathcal{M}_A = \{B \in \mathcal{M}(\mathcal{C}) \mid A \cap B \in \mathcal{M}(\mathcal{C})\}.$$

Let $C \in \mathcal{C}$, then $A \in \mathcal{M}(\mathcal{C}) = \mathcal{M}_C$
$= \{B \mid B \cap C \in \mathcal{M}(\mathcal{C})\}$

so that $A \cap C \in \mathcal{M}(\mathcal{C})$. Hence, $C \in \mathcal{M}_A$ and thus $\mathcal{C} \subseteq \mathcal{M}_A$.

The above proof for $A \in \mathcal{C}$ works also for

$A \in \mathcal{M}(\mathcal{C})$ so that $\mathcal{M}_A$ is a monotone class for any $A \in \mathcal{M}(\mathcal{C})$. This implies $\mathcal{M}(\mathcal{C}) \subseteq \mathcal{M}_A$ and hence $\mathcal{M}(\mathcal{C}) = \mathcal{M}_A$.

So, for any $A, B \in \mathcal{M}(\mathcal{C})$, $B \in \mathcal{M}_A$ so that $A \cap B \in \mathcal{M}(\mathcal{C})$. In particular, $\mathcal{M}(\mathcal{C})$ is closed under finite intersections.

This finishes the proof. ∎

Probability Measure on a Sigma Algebra:

Definition: Let $\mathcal{A}$ be a σ-algebra on a non-empty set S. A function $\mu : \mathcal{A} \to [0, \infty]$ is called a measure if it has the following properties:

1) $\mu(\phi) = 0$
2) $\mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$, where $A_k \in \mathcal{A}$ and

$A_k \cap A_\ell = \phi$ for all $k \neq \ell$.

In this case, the triple $(S, \mathcal{A}, \mu)$ is called a measure space.

Proposition: If A and B are in $\mathcal{A}$ and $A \subseteq B$ then $\mu(A) \leq \mu(B)$.

Proof: $B = A \cup (B \setminus A)$ is a disjoint union and hence, $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$, since $\mu(B \setminus A) \geq 0$. This finishes the proof. ∎

$\mu(S)$ is called the total mass of S. If $\mu(S) < +\infty$ then we say that the measure is finite. If $S = \bigcup_n A_n$, where $A_n \in \mathcal{A}$ for all n and $\mu(A_n) < +\infty$, then μ is called

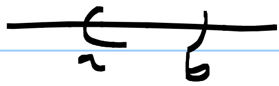
Video P

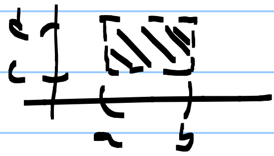
σ-finite.

<u>Examples:</u> 1) If S is a finite set and $\mathcal{A} = \mathcal{P}(S)$ let $\mu : \mathcal{P}(S) \longrightarrow [0,\infty]$ be the counting measure: $\mu(A) = |A|$ the number of elements in $A \in \mathcal{P}(S)$. It is clear that μ satisfies the measure axioms. In particular, this is a finite measure space.

2) Lebesgue measure μ is defined on Borel sets of $\mathbb{R}^d$ as follows:

$$\mu\left((a_1,b_1)\times\cdots\times(a_d,b_d)\right) = \prod_{i=1}^{d}(b_i - a_i).$$

<u>$d=1$</u>: $\mu((a,b)) = b-a$ 

<u>$d=2$</u>: $\mu((a,b)\times(c,d)) = (b-a)(d-c)$ 

Note that μ is <u>not</u> finite, however it is σ-finite:

$$\mathbb{R}^d = \bigcup_{n=1}^{\infty} \underbrace{(-n,n)\times\cdots\times(-n,n)}_{A_n}, \quad \mu(A_n) = (2n)^d$$

<u>Definition:</u> Let $(\Omega, \mathcal{A})$ be a measurable space. A measure $\mathbb{P}$ defined on the σ-algebra $\mathcal{A}$ of events, i.e., $\mathbb{P} : \mathcal{A} \longrightarrow \mathbb{R}$ satisfying the following condition $\mathbb{P}(\Omega) = 1$ is called a probability measure.

In particular a probability measure $\mathbb{P} : \mathcal{A} \to \mathbb{R}$ is characterized by the axioms:

P_1) $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$

P_2) $\forall A \in \mathcal{A}$, $\mathbb{P}(A) \geq 0$.

P_3) If $\{A_i \mid i \in I\}$ is a countable family of mutually exclusive events (pairwise disjoint sets) in $\mathcal{A}$, then

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathbb{P}(A_i).$$

Example: Let Ω be a set and $\mathcal{A}$ the family of all subsets of Ω. Fix some $\omega \in \Omega$ and define

$$\delta_\omega(A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases}$$

Clearly, δ_ω is a probability measure on $\mathcal{A}$, called the Dirac measure concentrated on ω.

Remark: Note a probability measure $\mathbb{P}$ on a σ-algebra $(\Omega, \mathcal{A})$ satisfies the followings:

1) If $A, B \in \mathcal{A}$ with $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.

2) If $A \in \mathcal{A}$ then $\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$.

3) If $A \in \mathcal{A}$ then
$$\begin{aligned} \mathbb{P}(A^c) &= \mathbb{P}(A) + \mathbb{P}(A^c) - \mathbb{P}(A) \\ &= \mathbb{P}(A \cup A^c) - \mathbb{P}(A) \\ &= 1 - \mathbb{P}(A). \end{aligned}$$

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

5) If $A, B \in \mathcal{A}$ then
$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \cup (B \setminus A)) \\ &= \mathbb{P}(A) + \mathbb{P}(B \setminus A) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B), \text{ since} \end{aligned}$$

$B = (B \setminus A) \cup (A \cap B)$ is a disjoint union.

6) If $A_1, \dots, A_n \in \mathcal{A}$ then

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} \mathbb{P}(A_k) - \sum_{1 \le k_1 < k_2 \le n} \mathbb{P}(A_{k_1} \cap A_{k_2})$$

$$+ \sum_{1 \le k_1 < k_2 < k_3 \le n} \mathbb{P}(A_{k_1} \cap A_{k_2} \cap A_{k_3})$$

$$\vdots$$

$$+ (-1)^{n-1} \mathbb{P}(A_1 \cap \cdots \cap A_n)$$

<u>Exercise</u>: Prove the above properties.

<u>Proposition</u>: (Monotone Sequential Continuity)

Let $\{A_n \mid n \in \mathbb{N}\}$ be a monotone sequence in $\mathcal{A}$. Then

$$\mathbb{P}(\lim_n A_n) = \lim_n \mathbb{P}(A_n)$$

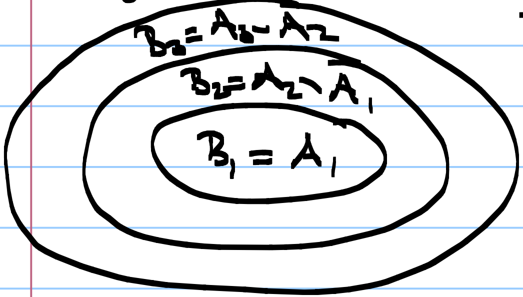
<u>Proof</u>: First let's take $\{A_n \mid n \in \mathbb{N}\}$ a nondecreasing sequence:

$$A_1 \subset A_2 \subset A_3 \subset \cdots\cdots$$

Then $\lim A_n = \bigcup_{n=1}^{\infty} A_n$. Note that

$$\lim A_n = \bigcup_{n=1}^{\infty} A_n = \bigcup_{k=1}^{\infty} (A_k \setminus A_{k-1}), \text{ where } A_0 = \emptyset, \text{ and}$$

$B_k \doteq A_k \setminus A_{k-1}$, $k = 1, 2, 3, \ldots$ is a pointwise disjoint sequence of events:



This implies that

$$p\left(\bigcup_{k=1} A_k\right) = p\left(\bigcup_{k=1}^{\infty} B_k\right)$$

$$= \sum_{k=1}^{\infty} p(B_k)$$

$$= \lim_{n \to \infty} \sum_{k=1}^{n} p(B_k)$$

$$= \lim_{n \to \infty} (p(B_1) + \cdots + p(B_n))$$

$$= \lim_{n\to\infty} \mu(B_1 \cup \cdots \cup B_n), \text{ since } B_i\text{'s are disjoint.}$$

$$= \lim_n \mu(A_n)$$

So, $\mu(\lim A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_n \mu(A_n)$.

I'll leave the "non-increasing case" as an exercise. ∎

Corollary (σ-subadditivity) Let $\{A_i \mid i \in I\}$ be any countable family of events in $\mathcal{A}$, then

$$\mathbb{P}\left(\bigcup_{i\in I} A_i\right) \leq \sum_{i \in I} \mathbb{P}(A_i).$$

Proof: Just take $I = \mathbb{N}$. Then $\bigcup_{i\in I} A_i = \bigcup_{n=1}^{\infty} A_n$.

Let $B_n \doteq \bigcup_{k=1}^{n} A_k$, for any $n \in \mathbb{N}$. Then $\{B_n \mid n \in \mathbb{N}\}$ is a monotone sequence ($B_n \subseteq B_{n+1}, \forall n$).
Now by the previous result applied to

$$\lim B_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty}\left(\bigcup_{k=1}^{n} A_k\right) = \bigcup_{n=1}^{\infty} A_n \text{ and obtain}$$

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \mathbb{P}(\lim B_n) = \lim \mathbb{P}(B_n).$$

Hence, $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_n \mathbb{P}(B_n)$

$$= \lim_n \mathbb{P}(A_1 \cup \cdots \cup A_n)$$

$$\leq \lim_n \sum_{i=1}^{n} \mathbb{P}(A_i)$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(A_n). \quad ∎$$

Video 9

Exercise: Let $\{A_i \mid i \in \mathbb{Z}\}$ and $\{B_i \mid i \in \mathbb{Z}\}$ be two countable families of events such that for all $i \in \mathbb{Z}$, we have $B_i \subseteq A_i$. Then

$$\mathbb{P}\Big(\bigcup_{i\in\mathbb{Z}} A_i\Big) - \mathbb{P}\Big(\bigcup_{i\in\mathbb{Z}} B_i\Big) \leq \sum_{i\in I} \big(\mathbb{P}(A_i) - \mathbb{P}(B_i)\big).$$

Corollary (Sequential Continuity)
Let $\{A_n \mid n \in \mathbb{N}\}$ be an arbitrary sequence of events, then

$$\mathbb{P}(\liminf A_n) \leq \liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n) \leq \mathbb{P}(\limsup A_n).$$

Proof: $\mathbb{P}(\liminf A_n) = \mathbb{P}\Big(\bigcup_{m=1}^{\infty} \bigcap_{n\geq m} A_n\Big)$

$$= \mathbb{P}\Big(\lim_m \Big(\bigcap_{n\geq m} A_n\Big)\Big)$$

$$= \lim_m \mathbb{P}\Big(\bigcap_{n\geq m} A_n\Big) \quad \text{(by the Monotone Seq. Continuity)}$$

$$= \lim_m \mathbb{P}\Big(\inf_{n\geq m} A_n\Big)$$

$\leq \liminf_m \mathbb{P}(A_m)$, where the last inequality follows since $\inf_{n\geq m} A_n \subseteq A_m$, so that $\mathbb{P}(\inf_{n\geq m} A_n) \leq \mathbb{P}(A_m)$ for all m and thus

$$\lim_m \mathbb{P}\Big(\inf_{n\geq m} A_n\Big) = \liminf \mathbb{P}\Big(\inf_{n\geq m} A_n\Big) \leq \liminf_m \mathbb{P}(A_m).$$

the sequence is monotone.

Similarly, $\mathbb{P}(\limsup_n A_n) \geq \limsup_m \mathbb{P}(A_m)$.

Finally, since $\liminf \mathbb{P}(A_n) \leq \limsup \mathbb{P}(A_n)$ we obtain the result.

Proposition: Let $\{A_n / n \in \mathbb{N}\} \subseteq \mathcal{A}$ such that $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty$. Then $\mathbb{P}(\{A_n, \text{ i.o.}\}) = 0$.

Proof: By σ-additivity

$$0 \leq \mathbb{P}\left(\bigcup_{n \geq m} A_n\right) \leq \sum_{n \geq m} \mathbb{P}(A_n).$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty$, $\lim_{m} \sum_{n \geq m} \mathbb{P}(A_n) = 0$.

Thus
$$0 = \lim_{m \to \infty} \sum_{n \geq m} \mathbb{P}(A_n) \geq \lim_{m \to \infty} \mathbb{P}\left(\bigcup_{n \geq m} A_n\right)$$
$$= \mathbb{P}\left(\lim_{m \to \infty} \bigcup_{n \geq m} A_n\right) \quad \left(\text{Monotone Seq. Cont.}\right)$$
$$= \mathbb{P}\left(\bigcap_m \bigcup_{n \geq m} A_n\right) \geq 0$$

Hence, $\mathbb{P}(\{A_n ; \text{ i.o.}\}) = \mathbb{P}(\limsup A_n) = \mathbb{P}\left(\bigcap_m \bigcup_{n \geq m} A_n\right) = 0$. ∎

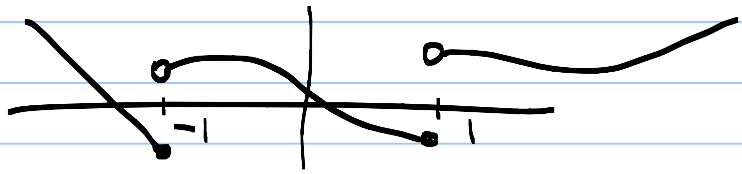
Definition: Let $(\Omega, \mathcal{A})$ be a measurable space and let $\mathbb{P}$ be a probability measure in $\mathcal{A}$. Then the triple $(\Omega, \mathcal{A}, \mathbb{P})$ is called a probability space.

Definition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A subset N of Ω is said to be negligible if there is a measurable set $A \in \mathcal{A}$ so that $N \subseteq A$ and $\mathbb{P}(A) = 0$. If $\mathcal{A}$ contains all the negligible sets then $(\Omega, \mathcal{A}, \mathbb{P})$ is said to be complete.

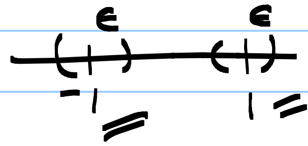
Definition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Then a property which holds for all $\omega \in \Omega$

except on a negligible subset of Ω is said to hold almost everywhere (a.e.) or $\mathbb{P}$-a.e. or a.a. $\omega \in \Omega$ or almost surely (a.s.) or $\mathbb{P}$-a.s.

<u>Example:</u>



This function is continuous on $\mathbb{R} \setminus \{-1, 1\}$. If $A = \{-1, 1\}$ then $\mu(A) = 0$. So f is continuous almost everywhere.



<u>Proposition:</u> If $\mathcal{N}$ denotes the class of all negligible sets of $(\Omega, \mathcal{A}, \mathbb{P})$ then $\bar{\mathcal{A}} = \{A \cup N \mid A \in \mathcal{A}, N \in \mathcal{N}\}$ coincides with the σ-algebra generated by $\mathcal{A}$ and $\mathcal{N}$. Moreover, the formula

$\bar{\mathbb{P}}(A \cup N) = \mathbb{P}(A)$ defines the unique probability $\bar{\mathbb{P}}$ on $\bar{\mathcal{A}}$, which extends $\mathbb{P}$.

<u>Example:</u> Consider the Lebesgue measure μ on $\mathbb{R}$.

$\mu(\{a\}) = 0$ since $\{a\} \subseteq (-\epsilon, \epsilon)$ for any $\epsilon > 0$ so that $0 \leq \mu(\{a\}) \leq \mu((-\epsilon, \epsilon)) = 2\epsilon$.

If $A = \{a_n \mid n \in \mathbb{N}\}$ then $A = \bigcup_{n \in \mathbb{N}} \{a_n\}$ and then

$$0 \leq \mu(A) \leq \sum_n \underbrace{\mu(\{a_n\})}_{0} = 0.$$

$A = \mathbb{Q}$ then $\mu(\mathbb{Q}) = 0$.

Video 10

Proof: Clearly, $\bar{\mathcal{A}}$ is contained in the σ-algebra generated by $\mathcal{A}$ and $\mathcal{N}$. Next, we'll show that $\bar{\mathcal{A}}$ is a σ-algebra, which will imply that $\bar{\mathcal{A}}$ contains the σ-algebra generated by $\mathcal{A}$ and $\mathcal{N}$, which would finish the proof of the first part.

Claim: $\bar{\mathcal{A}} = \{A \cup N \mid A \in \mathcal{A}, N \in \mathcal{N}\}$ is a σ-algebra.

Let $\{A_n \cup N_n \mid n \in \mathbb{N}\}$ be a countable class in $\bar{\mathcal{A}}$. Then

$\bigcup_{n=1}^{\infty}(A_n \cup N_n) = \left(\bigcup_{n=1}^{\infty} A_n\right) \cup \left(\bigcup_{n=1}^{\infty} N_n\right)$, where $\bigcup_{n=1}^{\infty} A_n$

is in $\mathcal{A}$ since $\mathcal{A}$ is a σ-algebra. To finish the proof we just need to show that $\bigcup_{n=1}^{\infty} N_n$ is a negligible set. Since each N_n is negligible there is a element $B_n \in \mathcal{A}$ so that $N_n \subseteq B_n$ and $\mathbb{P}(B_n) = 0$. Then $\bigcup_{n=1}^{\infty} N_n \subseteq \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$. Moreover,

$$0 \le \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \le \sum_{n=1}^{\infty} \mathbb{P}(B_n) = 0 \Rightarrow \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = 0.$$

Thus, $\bigcup_{n=1}^{\infty} N_n \in \mathcal{N}$ is a negligible set.

So, $\bigcup_{n=1}^{\infty}(A_n \cup N_n) \in \bar{\mathcal{A}}$.

To show that $\bar{\mathcal{A}}$ is an algebra we need to prove that $\bar{\mathcal{A}}$ is closed under taking complement. Let $A \in \mathcal{A}$, $N \in \mathcal{N}$. Choose $B \in \mathcal{A}$ with $N \subseteq B$ and $\mathbb{P}(B) = 0$.

$$\text{Then } (A \cup N)^c = (A \cup N)^c \cap (\overbrace{B \cup B^c}^{\Omega})$$
$$= [(A \cup N)^c \cap B] \cup [\underbrace{(A \cup N)^c \cap B^c}_{A^c \cap N^c}]$$

$$= [(A \cup N)^c \cap B] \cup [A^c \cap \underbrace{N^c \cap B^c}_{=B^c})$$

$$= [(A \cup N)^c \cap B] \cup (A^c \cap B^c)$$

Note that $A^c \cap B^c \in \mathcal{A}$ since A and B are in $\mathcal{A}$. Also, $(A \cup N)^c \cap B \subseteq B$ and $\mathbb{P}(B) = 0$ and thus $(A \cup N)^c \cap B \in \mathcal{N}$ is a negligible set. Hence, $(A \cup N)^c \in \overline{\mathcal{A}}$.

This concludes the proof that $\overline{\mathcal{A}}$ is a σ-algebra.

For the second part, first we must show that $\overline{\mathbb{P}}$ is well defined. For this let $A_1, A_2 \in \mathcal{A}$ and $N_1, N_2 \in \mathcal{N}$ so that

$$A_1 \cup N_1 = A_2 \cup N_2.$$

For $\overline{\mathbb{P}}(A_1 \cup N_1) = \overline{\mathbb{P}}(A_2 \cup N_2)$ we must show $\mathbb{P}(A_1) = \mathbb{P}(A_2)$.

Since $A_1 \cup N_1 = A_2 \cup N_2$ we have $(A_1 \setminus A_2) \cup (A_2 \setminus A_1) \subset N_1 \cup N_2$. Since

$(A_1 \setminus A_2) \cap (A_2 \setminus A_1) = \emptyset$ we obtain

$$0 \le \mathbb{P}(A_1 \setminus A_2) + \mathbb{P}(A_2 \setminus A_1) = \mathbb{P}((A_1 \setminus A_2) \cup (A_2 \setminus A_1)) = 0$$

because $(A_1 \setminus A_2) \cup (A_2 \setminus A_1) \subseteq N_1 \cup N_2$ and $N_1 \cup N_2$ is negligible.

In particular, $\mathbb{P}(A_1 \setminus A_2) = 0 = \mathbb{P}(A_2 \setminus A_1)$.

Finally,

$$A_1 = (A_1 \cap A_2) \cup (A_1 \setminus A_2) \text{ and hence}$$

$$\mathbb{P}(A_1) = \mathbb{P}(A_1 \cap A_2) + \underbrace{\mathbb{P}(A_1 \setminus A_2)}_{=0} = \mathbb{P}(A_1 \cap A_2).$$

Similarly, $\mathbb{P}(A_2) = \mathbb{P}(A_1 \cap A_2)$ and hence
$$\mathbb{P}(A_1) = \mathbb{P}(A_2).$$

Hence, $\overline{\mathbb{P}}$ is well defined. Next we need to prove the following:

i) $\overline{\mathbb{P}}(A \cup N) = \mathbb{P}(A) \geq 0$, for all $A \cup N \in \overline{\mathcal{A}}$.

ii) Let $\{A_n \cup N_n \mid n = 1, 2, \ldots\}$ be a disjoint class.
<u>must prove</u>:
$$\overline{\mathbb{P}}\left(\bigcup_{n=1}^{\infty}(A_n \cup N_n)\right) = \sum_{n=1}^{\infty} \overline{\mathbb{P}}(A_n \cup N_n)$$

This is left as an exercise. ∎

<u>Definition</u>: The probability space $(\Omega, \overline{\mathcal{A}}, \overline{\mathbb{P}})$ will be called the completion of $(\Omega, \mathcal{A}, \mathbb{P})$.

§1.3. <u>Interpretation of Probability and Classical Schemes</u>

Let Ω be a finite sample space, $\mathcal{A}$ a Boolean algebra of all subsets of Ω and $\mathbb{P}$ be a probability measure on $\mathcal{A}$. Assume that Ω has n elements and each element of Ω has the same probability $1/n$. If $A \in \mathcal{A}$ is an event then its probability is

$$\mathbb{P}(A) = \frac{1}{n} \cdot |A| = \frac{|A|}{|\Omega|}.$$

In probability theory, this is called the scheme of equally likely outcomes. Hence, to find the probability of a given event, it is important to count the number of elements it contains.

Video 11

Definition (Discrete Probabilities)
Let $A = \{\omega_i \mid i \in I\}$ be a countable subset of Ω and $\{p_i \mid i \in I\}$ a set of positive numbers such that $\sum_{i \in I} p_i = 1$. Then $\{p_i \mid i \in I\}$ determines a measure on the power set $\mathcal{P}(A)$ defined by

$$\mathbb{P}(B) = \sum_{\omega_i \in B} p_i, \quad B \in \mathcal{P}(A).$$

This clearly extends to any σ-algebra $\mathcal{A}$ of subsets of Ω containing $\mathcal{P}(A)$ by defining

$$\mathbb{P}(C) = \sum_{\omega_i \in C} p_i, \quad C \in \mathcal{A}.$$

Such a probability is called discrete.

Example: Let $\Omega = \mathbb{R}$, $A = \mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{P}(\{n\}) = (1/n^2)/(\pi^2/6)$.

If $B \subseteq \mathbb{R}$, $\mathbb{P}(B) = \mathbb{P}(B \cap \mathbb{N})$

Hence,
$$\begin{aligned}
\mathbb{P}(\mathbb{R}) &= \mathbb{P}(\mathbb{R} \cap \mathbb{N}) \\
&= \mathbb{P}(\mathbb{N}) \\
&= \frac{1}{\pi^2/6} \sum_{n=1}^{\infty} \mathbb{P}(\{n\}) \\
&= \frac{1}{\pi^2/6} \left(\sum_{n=1}^{\infty} \frac{1}{n^2} \right) \\
&= \frac{\pi^2/6}{\pi^2/6} = 1.
\end{aligned}$$

Geometric Probabilities:

Consider $\mathbb{R}^n$ together with its Borel σ-algebra and Lebesgue measure. Let $G \subseteq \mathbb{R}^n$ be a Borel set with finite Lebesgue measure $mes(G)$. Let $\mathcal{A}$ be the set of all Borel subsets of G and define $\mathbb{P}$ on $\mathcal{A}$ as follows:

$$\mathbb{P}(A) \doteq \frac{mes(A)}{mes(G)}, \quad A \in \mathcal{A}.$$

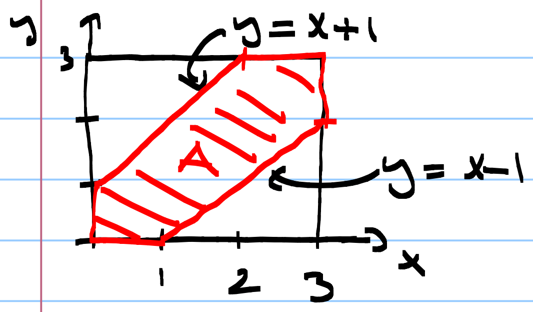
Exercise: Check that $\mathbb{P}$ is a probability measure on $\mathcal{A}$.

Example: (Problem of Meeting)
Two people agreed to meet between 12 a.m. and 3 p.m. They come to the place at random. The condition on this meeting problem is, the first coming person waits for one hour and then goes away. What is the probability of meeting?

Let $\Omega = \{(x,y) \in \mathbb{R}^2 \mid 0 \leq x,y \leq 3\}$ and let

$A = \{(x,y) \in \Omega \mid |x-y| \leq 1\}$.

Then we are asked $\mathbb{P}(A) = \dfrac{mes(A)}{mes(\Omega)} = \dfrac{5}{9}$

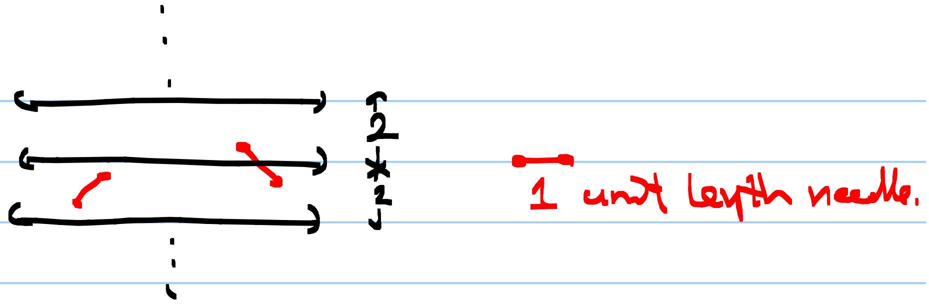


$|x-y| \leq 1$

$\iff -1 \leq y-x \leq 1$

$\iff x-1 \leq y \leq 1+x$

Example:



Compute the probability that a needle thrown into the air will touch one of lines on the ground.

§ 1.4. Conditional Probability and Independence:

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space fixed once and for all.

Conditional Probability: Let A and B be two events in $\mathcal{A}$ such that $\mathbb{P}(A) > 0$. Then the conditional probability of "B occuring under the condition that A is known to have occurred" (or conditional probability of B given A) denoted by $\mathbb{P}(B|A)$ is defined by

$$\mathbb{P}(B|A) \doteq \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)}.$$

Remarks: 1) In measure theory, as a function of B, the expression $\mathbb{P}(A\cap B)$ is called the restriction of $\mathbb{P}$ to A and $\mathbb{P}(A\cap B)/\mathbb{P}(A)$ is called the normalized restriction, since for $B = A$ we have $\mathbb{P}(A\cap A)/\mathbb{P}(A) = \mathbb{P}(A)/\mathbb{P}(A) = 1$.

2) If $\mathbb{P}(A) = 0$, then $\mathbb{P}(B|A)$ is defined to be zero for all B.

Example: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the mathematical model of tossing a die. Suppose we are in the classical scheme. Then, if it is given that an even number occurred in the die

i) What is the probability to have a 6?
ii) What is the probability to have a 5?

Solution: $A = \{2, 4, 6\}$, $B = \{6\}$, $C = \{5\}$.

i) $$\mathbb{P}(B/A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\{6\})}{\mathbb{P}(A)} = \frac{1/6}{3 \cdot 1/6} = 1/3.$$

ii) $$\mathbb{P}(C|A) = \frac{\mathbb{P}(C \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(A)} = \frac{0}{1/2} = 0.$$

Properties of the Conditional Probability

1) $\mathbb{P}(\cdot \mid A) : \mathcal{A} \to \mathbb{R}$ is also a probability measure on $\mathcal{A}$.

P_1) $$\mathbb{P}(\Omega, A) = \frac{\mathbb{P}(\Omega \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$$

P_2) $\forall B \in \mathcal{A}$, $$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \geq 0.$$

P_3) If $\{B_n \mid n \in \mathbb{N}\}$ is a countable collection of mutually disjoint events in $\mathcal{A}$, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \frac{\mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \cap A\right)}{\mathbb{P}(A)}$$
$$= \frac{\mathbb{P}\left(\bigcup_{n=1}^{\infty} (B_n \cap A)\right)}{\mathbb{P}(A)}$$

Since $B_n \cap B_m = \emptyset$ if $m \neq n$ so is $(B_n \cap A) \cap (B_m \cap A)$ and hence

$$\mathbb{P}\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \Big| A\right) = \frac{\sum_{n=1}^{\infty} \mathbb{P}(B_n \cap A)}{\mathbb{P}(A)}$$

$$= \sum_{n=1}^{\infty} \frac{\mathbb{P}(B_n \cap A)}{\mathbb{P}(A)}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}(B_n | A)$$

2) $\mathbb{P}(A|A) = \dfrac{\mathbb{P}(A \cap A)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$

3) If A and B are disjoint events, i.e., $A \cap B = \emptyset$, then $\mathbb{P}(B|A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(\emptyset)}{\mathbb{P}(A)} = 0.$

4) $$\mathbb{P}(B^c | A) = \frac{\mathbb{P}(B^c \cap A)}{\mathbb{P}(A)}$$

$$= \frac{\mathbb{P}(A \setminus (A \cap B))}{\mathbb{P}(A)}$$

$$= 1 - \mathbb{P}(B|A), \text{ because}$$

$A = (A \cap B) \cup (A \setminus (A \cap B))$, which is a disjoint union and thus

$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \setminus (A \cap B))$. Hence,

$1 = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} + \dfrac{\mathbb{P}(A \setminus (A \cap B))}{\mathbb{P}(A)}$ and therefore

$$\frac{\mathbb{P}(A \setminus (A \cap B))}{\mathbb{P}(A)} = 1 - \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = 1 - \mathbb{P}(B|A).$$

Video 12

Definition: The notion of conditional probability is extended to more than one event, as follows: Let $A_1, A_2, \dots, A_n$ be events in $\mathcal{A}$ such that $\mathbb{P}(A_1 \cap \dots \cap A_n) > 0$. Then the conditional probability of B given $A_1, \dots, A_n$ is defined as

$$\mathbb{P}(B \mid A_1, A_2, \dots, A_n) \doteq \frac{\mathbb{P}(B \cap A_1 \cap \dots \cap A_n)}{\mathbb{P}(A_1 \cap \dots \cap A_n)}.$$

Remark: If we have $\mathbb{P}(A_1 \cap \dots \cap A_k) > 0$ for all $k = 1, \dots, n$, then since

$$\mathbb{P}(A_n \mid A_1, \dots, A_{n-1}) = \frac{\mathbb{P}(A_1 \cap \dots \cap A_n)}{\mathbb{P}(A_1 \cap \dots \cap A_{n-1})}$$

we have

$$\begin{aligned}\mathbb{P}(A_1 \cap \dots \cap A_n) &= \mathbb{P}(A_n \mid A_1, \dots, A_{n-1})\, \mathbb{P}(A_1 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_n \mid A_1, \dots, A_{n-1})\, \mathbb{P}(A_{n-1} \mid A_1, \dots, A_{n-2}) \\ &\qquad \mathbb{P}(A_1 \cap \dots \cap A_{n-2}) \\ &= \mathbb{P}(A_1)\, \mathbb{P}(A_2 \mid A_1)\, \mathbb{P}(A_3 \mid A_1, A_2) \cdots \mathbb{P}(A_n \mid A_1, \dots, A_{n-1})\end{aligned}$$

Example: Suppose that the population of a certain city is 40% male and 60% female. Suppose also that 50% of males and 30% of the females smoke. Compute the probability that a smoker is male.

Solution: Let's define the following events.

M: the event a person selected is male

F: " " " " " " female

S: " " " " " smokes

N: " " " " " does not smoke

Then we have

$\mathbb{P}(S|M) = 0.5$, $\mathbb{P}(S|F) = 0.3$, $\mathbb{P}(M) = 0.4$ and $\mathbb{P}(F) = 0.6$

We are asked to determine $\mathbb{P}(M|S)$.

Then

$\mathbb{P}(M \cap S) = \mathbb{P}(S|M)\,\mathbb{P}(M) = 0.5 \times 0.4 = 0.2$ and $\mathbb{P}(S) = \mathbb{P}(S \cap M) + \mathbb{P}(S \cap F)$ since S is the disjoint union of $S \cap M$ and $S \cap F$. Also,

$$\mathbb{P}(S \cap F) = \mathbb{P}(F)\,\mathbb{P}(S|F) = 0.6 \times 0.3 = 0.18$$

So, $\mathbb{P}(S) = \mathbb{P}(S \cap M) + \mathbb{P}(S \cap F) = 0.2 + 0.18 = 0.38$

Thus, $\mathbb{P}(M|S) = \dfrac{\mathbb{P}(M \cap S)}{\mathbb{P}(S)} = \dfrac{0.2}{0.38} \simeq 0.53$

This is a special case of so called

Bayes Formula: Let $\{B_1, \dots, B_n\}$ be a finite partition of Ω. Let A be an event in $\mathcal{A}$ such that $\mathbb{P}(A) > 0$ and suppose that $\mathbb{P}(A|B_j)$ and $\mathbb{P}(B_j)$ are specified. Then we may compute $\mathbb{P}(B_i|A)$ for $i = 1, \dots, n$, as follows:

$$A = A \cap \Omega = A \cap (B_1 \cup \dots \cup B_n) = \bigcup_{j=1}^{n} (A \cap B_j).$$

Then $\mathbb{P}(A) = \sum_{j=1}^{n} \mathbb{P}(A \cap B_j)$

$$= \sum_{j=1}^{n} \mathbb{P}(A|B_j)\,\mathbb{P}(B_j) \quad \text{(Theorem of Total Probability)}$$

So, $\mathbb{P}(B_i|A) = \dfrac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(B_i)\,\mathbb{P}(A|B_i)}{\sum_{j=1}^{n} \mathbb{P}(B_j)\,\mathbb{P}(A|B_j)}$ (Bayes Formula)

Example: There are two urns U_1 and U_2 that contain 2 white and 2 black balls and 2 white and 3 black balls, respectively. A ball is chosen from U_1 and transferred into U_2. Then a ball is chosen from U_2 and it turns out to be black. What is the probability that the first ball chosen from U_1 was white?

Solution: Define the following events:

B_1: the ball drawn from U_1 is white

B_2: the ball drawn from U_1 is black, and

A: the ball drawn from U_2 is black.

We want to compute $\mathbb{P}(B_1 \mid A)$. $\Omega = B_1 \cup B_2$

$$\mathbb{P}(B_1 \mid A) = \frac{\mathbb{P}(B_1)\,\mathbb{P}(A \mid B_1)}{\mathbb{P}(B_1)\,\mathbb{P}(A \mid B_1) + \mathbb{P}(B_2)\,\mathbb{P}(A \mid B_2)}$$

Note that $\mathbb{P}(B_1) = 1/2$, $\mathbb{P}(B_2) = 1/2$

$\mathbb{P}(A \mid B_1) = 1/2$, $\mathbb{P}(A \mid B_2) = 4/6$.

So, $$\mathbb{P}(B_1 \mid A) = \frac{1/2 \cdot 1/2}{1/2 \cdot 1/2 + 1/2 \cdot 2/3} = \frac{1/4}{1/4 + 1/3} = \frac{3}{7}.$$

Video 13

Example (Total Probability)
Let's assume that k urns are given: $U_1, \ldots, U_k$, where U_j contains m_j white and n_j black balls, $j=1,\ldots,k$. Take at random an urn. What is the probability to draw one white ball from that urn?

Solution: Let A be the event of taking one white ball and B_i be the event of taking the i^{th} urn.

Then by the Theorem of Total Probability we have

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A|B_i)\, \mathbb{P}(B_i)$$

$$= \sum_{i=1}^{k} \frac{m_i}{m_i+n_i} \times \frac{1}{k}$$

$$= \frac{1}{k} \sum_{i=1}^{k} \frac{m_i}{m_i+n_i}$$

Example: Suppose that there are three chests A_1, A_2, A_3 each having two drawers, so that one contains one gold in each drawer, one contains one silver coin in each drawer and one contains one silver in one drawer and one gold coin the other drawer. Let A_1, A_2, A_3 denote the events the above described chests are chosen, respectively.

A chest is chosen at random and a drawer is opened. If the drawer contains a gold coin, what is the probability that the other drawer also contains a gold coin?

Solution: We may regard the events A_1, A_2 and A_3 as disjoint events and $\Omega = A_1 \cup A_2 \cup A_3$. Therefore

$\mathbb{P}(A_i) = 1/3$, $i=1,2,3$. Let B be the event that the coin observed was gold. Then we have

$$\mathbb{P}(B \mid A_1) = 1, \quad \mathbb{P}(B \mid A_2) = 1/2 \text{ and } \mathbb{P}(B \mid A_3) = 0$$

The problem asks for the probability that the second drawer has a gold coin given that there was a coin in the first. This is possible only if the chest chosen is the first one. Hence, the problem is equivalent to compute $\mathbb{P}(A_1 \mid B)$. By the Bayes Formula

$$\mathbb{P}(A_1 \mid B) = \frac{\mathbb{P}(A_1)\,\mathbb{P}(B \mid A_1)}{\sum_i \mathbb{P}(A_i)\,\mathbb{P}(B \mid A_i)} = \frac{1/3 \times 1}{1/3 \times 1 + 1/3 \times 1/2 + 1/3 \times 0} = \frac{2}{3}.$$

<u>Exercise</u>: Compute the probability that the second drawer has a silver coin given that the first drawer had a gold coin?

<u>Independence of Events</u>:

<u>Definition</u>: Two events A and B are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

<u>Example</u>: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the probability space of tossing two die. Assume the classical rules that $\mathbb{P}(A) = \dfrac{|A|}{|\Omega|}$, for all $A \in \mathcal{A}$.

Let $A_1 = \{1 \text{ appears on the } 1^{st} \text{ die}\}$ and
$A_2 = \{2 \text{ appears on the } 2^{nd} \text{ die}\}$.

$$\mathbb{P}(A_1) = \frac{|A_1|}{|\Omega|} = \frac{6}{36} = 1/6$$

$$\mathbb{P}(A_2) = \frac{|A_2|}{|\Omega|} = \frac{6}{36} = 1/6.$$

$$\mathbb{P}(A_1 \cap A_2) = \frac{|A_1 \cap A_2|}{|\Omega|} = \frac{1}{36} = \mathbb{P}(A_1)\,\mathbb{P}(A_2).$$

Hence, the events A_1 and A_2 are independent.

<u>Lemma</u>: If A and B are independent events in $\mathcal{A}$, then so are A^c, B and A, B^c and A^c, B^c.

<u>Proof</u>: Let's prove only the first one.

$$\mathbb{P}(A) = \mathbb{P}(A \cap \underbrace{(B \cup B^c)}_{=\Omega}) = \mathbb{P}((A \cap B) \cup (A \cap B^c))$$

$\Rightarrow \mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$ since the events are disjoint

$$\begin{aligned}
\Rightarrow \mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A)\,\mathbb{P}(B) \\
&= \mathbb{P}(A)\,(1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A)\,\mathbb{P}(B^c)
\end{aligned}$$

Hence, A and B^c are independent. ∎

<u>Remarks</u>: 1) Ω is independent from any event:

$$\mathbb{P}(A \cap \Omega) = \mathbb{P}(A) \cdot 1 = \mathbb{P}(A)\,\mathbb{P}(\Omega).$$

2) $\emptyset$ is independent from any event:

$$\mathbb{P}(A \cap \emptyset) = \mathbb{P}(\emptyset) = 0 = \mathbb{P}(A) \cdot 0 = \mathbb{P}(A)\,\mathbb{P}(\emptyset)$$

3) If A is an event independent from itself then $\mathbb{P}(A) = 0$ or 1:

Video 14

$$\mathbb{P}(A \cap A) = \mathbb{P}(A)\,\mathbb{P}(A)$$
$\quad \parallel$

$\mathbb{P}(A)$

$$\Rightarrow \mathbb{P}(A) - \mathbb{P}(A)\,\mathbb{P}(A) = 0$$
$$\Rightarrow \mathbb{P}(A)\,(1 - \mathbb{P}(A)) = 0$$

Then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

__Definition:__ Given events $A_1, \dots, A_n$ are said to be mutually independent if for every possible subset of k $(2 \le k \le n)$ of the events $A_1, \dots, A_n$ we have

$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$, where

$1 \le i_1 < i_2 < \dots < i_k \le n$.

__Remark:__ Let $n = 3$ and assume that A_1, A_2 and A_3 are mutually independent. Then we have the following conditions:

$\mathbb{P}(A_i) = \mathbb{P}(A_i) \quad i = 1, 2, 3$

$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j)$, $1 \le i < j \le 3$, and

$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\,\mathbb{P}(A_3)$.

Note that the equations above are all independent in the sense that none of them can be derived from the rest.

__Proposition:__ The events $A_1, A_2, \dots, A_n$ are mutually independent if and only if

$$\mathbb{P}(A_1^{\alpha_1} \cap \dots \cap A_n^{\alpha_n}) = \mathbb{P}(A_1^{\alpha_1})\,\mathbb{P}(A_2^{\alpha_2}) \cdots \mathbb{P}(A_n^{\alpha_n}), \text{ when}$$

$A_i^{\alpha_i} = A_i$ or A_i^c, for each $i = 1, \dots, n$.

__Proof__ is left as an exercise.

__Remark:__ Let $\mathcal{C} = \{A_1, A_2, \dots, A_n\}$ be a finite

collection of events so that $A_1 \cap A_2 \cap \cdots \cap A_n = \emptyset$. Assume that each $\mathbb{P}(A_i) > 0$. Then $\mathcal{C}$ is not a collection of mutually independent events because
$$0 = \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2)\cdots\mathbb{P}(A_n) > 0,$$
a clear contradiction.

Definition. The events $\{A_1, A_2, \ldots, A_n\}$ are said to be pairwise independent if each pair of events A_i and A_j are independent for any $1 \leq i < j \leq n$.

Remark: Pairwise independence does not imply mutual independence: $\Omega = \{(0,0,0), (0,1,1), (1,0,1), (1,1,0)\}$ and define the events

$A = \{$the first digit is zero$\}$,
$B = \{$the second digit is zero$\}$,
$C = \{$the third digit is zero$\}$.

Then, we have $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = 1/2$, $\mathbb{P}(C) = \frac{1}{2}$.

$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap C) = \mathbb{P}(A \cap C) = \frac{1}{4}$.

$\mathbb{P}(A)\mathbb{P}(B) = 1/4$, $\mathbb{P}(A)\mathbb{P}(C) = 1/4$, $\mathbb{P}(B)\mathbb{P}(C) = 1/4$

Hence, these events are pairwise independent.

$\mathbb{P}(A \cap B \cap C^c) = 0 \neq \frac{1}{8} = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C^c)$ so that

A, B and C are not mutually independent.

Definition. Let $A \in \mathcal{A}$ have positive probability and let $\mathbb{P}(\cdot \mid A)$ be the conditional probability on $\mathcal{A}$ given A. A class $\{B_1, B_2, \ldots, B_n\}$ in $\mathcal{A}$ is said to be conditionally mutually independent given A, if it is mutually independent with respect to the conditional probability $\mathbb{P}(\cdot \mid A)$.

In particular, two events B and C are conditionally independent given A if

$$\mathbb{P}(B \cap C \mid A) = \mathbb{P}(B \mid A)\, \mathbb{P}(C \mid A).$$

Note that in this case,

$$\frac{\mathbb{P}(B \cap C \cap A)}{\mathbb{P}(A)} = \mathbb{P}(B \mid A)\, \mathbb{P}(C \mid A) \quad \text{so that}$$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\, \mathbb{P}(B \mid A)\, \mathbb{P}(C \mid A).$$

<u>Remark</u>: Note that $\mathbb{P}(C \mid A \cap B) = \dfrac{\mathbb{P}(C \cap A \cap B)}{\mathbb{P}(A \cap B)}$

$$= \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(A)\, \mathbb{P}(B \mid A)}$$

and thus

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\, \mathbb{P}(B \mid A)\, \underbrace{\mathbb{P}(C \mid A \cap B)}_{\mathbb{P}(C \mid A, B)}.$$

<u>Exercise</u>: Using Induction prove that

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_k) = \mathbb{P}(A_1)\, \mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_k \mid A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

<u>Proposition</u>: B and C are independent given A if and only if $\mathbb{P}(C \mid A, B) = \mathbb{P}(C \mid A)$ or equivalently $\mathbb{P}(B \mid A, C) = \mathbb{P}(B \mid A)$, where we assume that the defining probabilities are positive.

Proof: $\mathbb{P}(\underline{B\cap C}|A) = \dfrac{\mathbb{P}(A\cap B\cap C)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(A)\,\mathbb{P}(B|A)\,\mathbb{P}(C|A,B)}{\mathbb{P}(A)}$

$$= \mathbb{P}(B|A)\,\mathbb{P}(C|A,B).$$

So, if B and C are independent given A, then

$\mathbb{P}(B|A)\,\mathbb{P}(C|A) = \mathbb{P}(B\cap C|A)$ so that

$\mathbb{P}(C|A) = \mathbb{P}(C|A,B)$.

Conversely, if $\mathbb{P}(C|A) = \mathbb{P}(C|A,B)$ then going backwards we see that
$\mathbb{P}(B\cap C|A) = \mathbb{P}(B|A)\,\mathbb{P}(C|A)$ so that B and C are independent given A.
The other statement is proved similarly.

Definition: A finite collection $\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n\}$ of sub-σ-algebras of $\mathcal{A}$ is said to be (mutually) independent (with respect to a probability $\mathbb{P}$ on $\mathcal{A}$) if

$$\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big) = \prod_{i=1}^{n} \mathbb{P}(A_i), \text{ for all } A_i \in \mathcal{A}_i.$$

An infinite family of sub-σ-algebras of $\mathcal{A}$ is said to be independent if every finite subfamily is independent.

Video 15

<u>Proposition</u>: Let $\{A_1, \dots, A_n\}$ be a collection of events in $\mathcal{A}$. Let $\mathcal{A}_i = \{\phi, A_i, A_i^c, \Omega\}$ be the sub-Boolean algebra of $\mathcal{A}$ generated by A_i. Then the events $\{A_1, \dots, A_n\}$ are mutually independent if and only if the corresponding sub-Boolean algebras $\{\mathcal{A}_1, \dots, \mathcal{A}_n\}$ are mutually independent.

Proof is immediate by the definition and previous propositions.

<u>Proposition</u>: A collection $\{\mathcal{A}_1, \dots, \mathcal{A}_n\}$ of finite sub-Boolean algebras of $\mathcal{A}$ is mutually independent if and only if

$$(*) \quad \mathbb{P}\left(\bigcap_{j=1}^{n} A_l^j\right) = \prod_{j=1}^{n} \mathbb{P}(A_l^j), \text{ where } A_l^j \text{ is an}$$

arbitrary atom of $\mathcal{A}_j$, with $1 \le l \le m_j$, where m_j is the number of atoms in $\mathcal{A}_j$ and $1 \le j \le n$.

<u>Proof</u>: The "$\Rightarrow$" direction is clear by definition.

"$\Leftarrow$": Now suppose that the equation $(*)$ holds for atoms. Let $\{A_1^j, \dots, A_{m_j}^j\}$ be the set of all atoms of $\mathcal{A}_j$, $j = 1, \dots, n$. Then an element D_j of $\mathcal{A}_j$ can be written as

$$D_j = \bigcup_{j=1}^{m_j} (\alpha_l^j \cap A_l^j), \text{ where } \alpha_l^j = \phi \text{ or } \Omega.$$

So we need to prove $\mathbb{P}\left(\bigcap_{j=1}^{n} D_j\right) = \prod_{j=1}^{n} \mathbb{P}(D_j)$.

Let $\alpha_l^j \cap A_l^j = B_l^j$ for $l = 1, 2, \dots, m_j$, $j = 1, 2, \dots, n$.

Then $\bigcap_{j=1}^{n} D_j = \bigcap_{j=1}^{n} \left(\bigcup_{j=1}^{m_j} B_l^j\right) = \bigcup_{k} \bigcap_{j=1}^{n} B_{l_k}^j$, where

$(l_1, \dots, l_n)$ is any point of $\{1,2,\dots,m_1\} \times \{1,2,\dots,m_2\} \times \dots \times \{1,2,\dots,m_n\}$ and the union is taken over all the points of the product set.

The class $\{\bigcap_{j=1}^{n} B^j_{l_k}, l_k\}$ is a disjoint class.

Thus, $\mathbb{P}(\bigcap_{j=1}^{n} D_j) = \sum_{l_k} \mathbb{P}(\bigcap_{j=1}^{n} B^j_{l_k})$

$$= \sum_{l_k} \prod_{j=1}^{n} \mathbb{P}(B^j_{l_k}) \quad \text{(by the assumption that "atoms" are independent)}$$

$$= \prod_{j=1}^{n} \sum_{j=1}^{m_j} \mathbb{P}(B^j_{l_k})$$

$$= \prod_{j=1}^{n} \mathbb{P}(D_j).$$

Hence, $\{A_1, \dots, A_n\}$ is mutually independent ∎

<u>Corollary</u> Let $\mathcal{C} = \{A_1, \dots, A_n\}$ be a class of mutually independent events. Let $\{\mathcal{C}_{k_1}, \mathcal{C}_{k_2}, \dots, \mathcal{C}_{k_l}\}$ be a disjoint subclass of $\mathcal{C}$. Let $\mathcal{A}_{k_1}, \mathcal{A}_{k_2}, \dots, \mathcal{A}_{k_l}$ be the sub-Boolean algebras generated by $\mathcal{C}_{k_1}, \mathcal{C}_{k_2}, \dots, \mathcal{C}_{k_l}$, respectively. Then $\mathcal{A}_{k_1}, \dots, \mathcal{A}_{k_l}$ are mutually independent.

<u>Example:</u> Let $\{A, B, C, D, E\}$ be mutually independent class of events. We claim that A, $B \cup C$, $D \cup E^c$ are mutually independent.

<u>Proof:</u> Consider the classes $\mathcal{C}_{k_1} = \{A\}$, $\mathcal{C}_{k_2} = \{B, C\}$ and $\mathcal{C}_{k_3} = \{D, E\}$. Define $\mathcal{A}_{k_r} = \sigma(\mathcal{C}_{k_r})$, $r = 1, 2, 3$.

By the corollary, $\mathcal{A}_{k_i}$'s are mutually independent. In particular, A, $B \cup C$, $D \cup E^c$ are mutually independent.

<u>Proposition</u>: (Criterion for Independence)
Let $\{\mathcal{C}_i \mid i \in I\}$ be an arbitrary family of non empty subclass of the σ-algebra $\mathcal{A}$ having the following properties:

i) each $\mathcal{C}_i$ is closed under intersection.
ii) $\{\mathcal{C}_i \mid i \in I\}$ is independent, in the sense that for every choice of $C_j \in \mathcal{C}_j$, $j \in J \subseteq I$, with finite J we have

$$\mathbb{P}\Big(\bigcap_{j \in J} C_j\Big) = \prod_{j \in J} \mathbb{P}(C_j).$$

Then $\{\sigma(\mathcal{C}_i) \mid i \in I\}$ is independent and every family $\{\mathcal{F}_i \mid i \in I\}$ of σ-algebras differing from $\{\sigma(\mathcal{C}_i) \mid i \in I\}$ only on negligible sets is also independent.

<u>Proof</u>: By the definition it is enough to check independence for only finite index sets I. So let $\mathcal{C}_1, \dots, \mathcal{C}_n$ be subclasses (of $\mathcal{A}$) having the properties listed in the proposition. Let $\mathcal{D}_1$ be the class of all elements $D_1 \in \mathcal{A}_1$ satisfying

$$\mathbb{P}\Big(D_1 \cap \Big(\bigcap_{j=2}^{n} C_j\Big)\Big) = \mathbb{P}(D_1) \prod_{j=2}^{n} \mathbb{P}(C_j),$$

for any choice of $C_j \in \mathcal{C}_j$, $j = 2, \dots, n$. By the first lemma of this subsection, if $D_1 \in \mathcal{D}_1$ then $D_1^c \in \mathcal{D}_1$. Next consider the class $\mathcal{B}_1$ consisting of elements D_1 such that $D_1 \cap C_1 \in \mathcal{C}_1$.

Let $D_1 \in \mathcal{B}_1$ and $C_1 \in \mathcal{C}_1$. Then $D_1 \cap C_1 \in \mathcal{C}_1$. Let $K = \bigcap_{j=2}^{n} C_j$. Then we have

$$\begin{aligned}\mathbb{P}(D_1^c \cap C_1 \cap K) &= \mathbb{P}(C_1 \cap K) - \mathbb{P}(D_1 \cap C_1 \cap K)\\ &= \mathbb{P}(C_1)\,\mathbb{P}(K) - \mathbb{P}(D_1 \cap C_1)\,\mathbb{P}(K)\\ &= [\mathbb{P}(C_1) - \mathbb{P}(D_1 \cap C_1)]\,\mathbb{P}(K)\\ &= \mathbb{P}(C_1 \cap D_1^c)\,\mathbb{P}(K).\end{aligned}$$

Hence, $C_1 \cap D_1^c \in \mathcal{C}_1$ implies $D_1^c \in \mathcal{B}_1$. Now we claim that $\mathcal{B}_1$ is closed under intersection: let $D_1, D_2 \in \mathcal{B}_1$. Since $D_1 \in \mathcal{B}_1$, $D_1 \cap C_1 \in \mathcal{C}_1$ and $D_1^c \cap C_1 \in \mathcal{C}_1$. Similarly, $D_2 \cap C_1$ and $D_2^c \cap C_1$ both belong to $\mathcal{C}_1$. These imply $(D_1 \cap D_2) \cap C_1 = D_1 \cap (D_2 \cap C_1) \in \mathcal{C}$ and thus $D_1 \cap D_2 \in \mathcal{B}_1$. Since $C_1 \cap C_2 \in \mathcal{C}_1$, $\mathcal{C}$ is contained in $\mathcal{B}_1$. Hence $\mathcal{B}_1$ is a Boolean algebra containing $\mathcal{C}_1$ and included in $\mathcal{D}_1$, which is a monotone class, i.e. $\mathcal{C}_1 \subset \mathcal{B}_1 \subset \mathcal{D}_1$. By the Monotone Class Theorem for Sets, we see that $\mathcal{D}_1$ contains $\sigma(\mathcal{B}_1)$ and hence $\sigma(\mathcal{C}_1)$. Now we see that the family $\{\sigma(\mathcal{C}_1), \mathcal{C}_2, \ldots, \mathcal{C}_n\}$ has the two properties of the proposition. Continuing this way we can prove that $\sigma(\mathcal{C}_1), \sigma(\mathcal{C}_2), \ldots, \sigma(\mathcal{C}_n)$ form an independent family.

Finally, if the σ-algebras $\mathcal{F}_i$ differ from $\sigma(\mathcal{C}_i)$ by negligible sets, the fact that sets differing by negligible sets have the same probability, which implies the independence of $\{\mathcal{F}_i, i \in I\}$. ∎

Theorem (Borel–Cantelli Lemma)

Let $\{A_n \mid n \in \mathbb{N}\} \subseteq \mathcal{A}$. Then

i) $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < +\infty \Rightarrow \mathbb{P}(\{A_n, i.o.\}) = 0.$

ii) If $\{A_n \mid n \in \mathbb{N}\}$ is independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = +\infty$

then $\mathbb{P}(\{A_n, i.o.\}) = 1.$

Proof: (i) is already proved

ii) So we assume that $\{A_n \mid n \in \mathbb{N}\}$ is independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty.$

Then
$$\mathbb{P}\Big(\bigcup_{n \le k \le N} A_k\Big) = 1 - \mathbb{P}\Big[\Big(\bigcup_{n \le k \le N} A_k\Big)^c\Big]$$
$$= 1 - \mathbb{P}\Big(\bigcap_{n \le k \le N} A_k^c\Big)$$
$$= 1 - \prod_{n \le k \le N} \mathbb{P}(A_k^c) \quad \text{(since } \{A_k, A_k^c \mid k \in \mathbb{N}\} \text{ is independent)}$$
$$= 1 - \prod_{n \le k \le N} (1 - \underbrace{\mathbb{P}(A_k)}_{x})$$

Now we have the following fact: $1 - x \le e^{-x}$, for all $x \ge 0$.

Proof: Let $f(x) = e^{-x} + x - 1$. Then $f(0) = 0$, $f'(x) = -e^{-x} + 1$ so that $f'(0) = 0$ and $f''(x) = e^{-x} > 0$, for all $x \ge 0$. Hence $f'(x)$ is strictly increasing and thus $f'(x) > 0$ for $x > 0$. In particular, $f'(x) \ge 0$ so that f is increasing. Since $f(0) = 0$, $f(x) \ge 0$ if $x \ge 0$.

Now, $\mathbb{P}\left(\bigcup_{n\le k\le N} A_k\right) = 1 - \prod_{n\le k\le N} (1-\mathbb{P}(A_k))$

$$\geq 1 - \prod_{n\le k\le N} e^{-\mathbb{P}(A_k)}$$

$$= 1 - e^{-\sum_{n\le k\le N}\mathbb{P}(A_k)}$$

Also note that, by the assumption $\lim_{N\to\infty} \sum_{k=n}^{N} \mathbb{P}(A_n) = \infty$

and hence $\mathbb{P}\left(\bigcup_{n\le k\le N} A_k\right) \to 1$ as $N \to \infty$.

Then $\mathbb{P}(\{A_n, \text{i.o.}\}) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\ge n} A_k\right) = 1.$ ∎

<u>Example:</u> What is the probability that in a sequence of Bernoulli trials the pattern SFS (Success – Failure) appears infinitely often?

<u>Solution:</u> Let A_k be the event that the trials number k, $k+1$ and $k+2$ produce the sequence SFS. The events A_k's are not mutually independent. However, the sequence of events $A_1, A_4, A_7, \dots, A_{3k+1}, \dots$ consists of mutually independent events, because no two of them depend on the outcome of the same trial.

Since $\mathbb{P}(A_k) = a_k = p^2q$, for some $p,q>0$ with $p+q=1$, the series

$$a_1+a_4+a_7+a_{10}+\dots\dots = \sum_{k=0}^{\infty} a_{3k+1} = \sum_{k=0}^{\infty} p^2q = +\infty$$

and hence the pattern SFS occurs infinitely often with probability 1.

Video 17 a,b

CHAPTER 2: Combinatorial Problem and Equally Likely Outcomes:

§2.1. Counting Principles:

Definition: Let $A=\{a_1, a_2, \dots, a_n\}$ be an alphabet of n letters. Any finite succession of letters will be called a word (meaningful or not). The number of letters in a word will be called the length of the word. Given a word $A = a_{j_1} a_{j_2} \dots a_{j_k}$ of length k, the first succession of l $(1 \le l \le k)$ letters $a_{j_1} a_{j_2} \dots a_{j_l}$, will be called the initial part of A of length l. A word that has no letters will be called the empty word and will be denoted by Λ. The length of Λ is zero.

Two kinds of words will be considered: H_1 will be the class of words in which every letter can be found at most once. Clearly, in this case, the length of any word will not exceed n, the number of letters in the alphabet. H_2 will be the class of words, where the letters of the alphabet may appear more than once. The length of such word can be, of course, arbitrarily large.

Proposition: Let $A = a_{j_1} a_{j_2} \dots a_{j_k}$ be a word of length k in H_1. Then there are $(n-k)$ different words of length $k+1$ in H_1 having A as initial expression.

More generally there are $(n-k)(n-k-1)\dots(n-j+1)$ different words in H_1 of length $j > k$, having A as initial expression.

Proof:

$$\overbrace{\underbrace{a_{j_1} a_{j_2} \cdots a_{j_k}}_{k}}\ \underbrace{\overset{n-k \text{ choices}}{?}}_{1} \in H_1$$

n−k choices, $n-(k+1) = (n-k-1)$ choices, $n-j+1$ choices

$$\overbrace{\underbrace{a_{j_1} a_{j_2} \cdots a_{j_k}}_{k}\ \underbrace{\ast\ \ast \cdots \ast}_{j-k}}^{j}$$

Hence, $(n-k)\cdot(n-k-1)\cdots(n-j+1)$ is the total number of choices. ∎

Corollary The number of different words of length ℓ in H_1 is equal

$$n^{(\ell)} = n(n-1)\cdots(n-\ell+1).$$

Definition: Let $\mathcal{A}_n$ be the set of all words of length n in H_1. Let $\mathcal{P} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_i\}$ be a partition of alphabet $\mathcal{A}$. ($\mathcal{A} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \dots \cup \mathcal{C}_i$ is a disjoint union.) Two words A_1 and A_2 in $\mathcal{A}_n$ are said to be $\mathcal{P}$-equivalent iff, for all $j = 1, \dots, n$, the jth letter of A_1 and A_2 belong to the same $\mathcal{C}_h$ $(h = 1, 2, \dots, i)$. The $\mathcal{P}$-equivalence class of a word $A \in \mathcal{A}_n$ is the set of all words in $\mathcal{A}_n$ which are equivalent to A.

Example: Let $\mathcal{A} = \{a, b, c, d, e\}$ be an alphabet and let $\mathcal{C}_1 = \{b, c, d\}$ and $\mathcal{C}_2 = \{a, e\}$ be a partition $\mathcal{P}$ of $\mathcal{A}$. Then the words abcda and ecbda are $\mathcal{P}$-equivalent.

Proposition: Let $\mathcal{A}_n$ and $\mathcal{P}$ be as in the above definition and let $k_1, k_2, \dots, k_i$ be the number of letters in $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_i$, respectively. Then the number of different $\mathcal{P}$-equivalence classes of $\mathcal{A}_n$ is

$$\binom{n}{k_1, k_2, \ldots, k_{i-1}} = \frac{n!}{k_1!\, k_2! \cdots k_{i-1}!\, k_i!}, \text{ where}$$

$k_1 + k_2 + \ldots + k_i = n$.

Proof: Note that each $\mathcal{P}$-equivalence class has $k_1!\, k_2! \ldots k_i!$ words. Since A_n has $n!$ elements the number of $\mathcal{P}$-equivalence classes is

$$n! / k_1!\, k_2! \cdots k_i!.$$

Corollary: Let $\mathcal{C} \subseteq A$ and $\mathcal{P} = \{\mathcal{C}, \mathcal{C}^c\}$. If $|\mathcal{C}| = k$ then the number of $\mathcal{P}$-equivalence classes in A_n is

$$\frac{n!}{k!\,(n-k)!} = \binom{n}{k}.$$

Proposition: Let $A \in H_2$ have length k. Then there are n different words of length $k+1$ in H_2 having A as its initial expression.

In particular the number of words of length k in H_2 is n^k.

Proposition: The number of different words of length N in H_2 in which all the letters $a_1, a_2, \ldots, a_n$ occurs $k_1, k_2, \ldots, k_n$ times, respectively, equals

$$\binom{N}{k_1, k_2, \ldots, k_{n-1}} = \frac{N!}{k_1!\, k_2! \cdots k_n!}, \text{ when } k_1 + \ldots + k_n = N.$$

Proof Consider a new alphabet

$$A' = \{a_1^1, a_1^2, \ldots, a_1^{k_1}, a_2^1, a_2^2, \ldots, a_2^{k_2}, \ldots, a_n^1, a_n^2, \ldots, a_n^{k_n}\}$$

with $k_1+k_2+\cdots+k_n=N$. Note that two words of length N, when each letter of A' is used only once, are equivalent if they generate the same word with repeated letters in A, when the upper indices are erased.

Now let $C_1=\{a_1^1,a_1^2,\ldots,a_1^{k_1}\}$, $C_2=\{a_2^1,a_2^2,\ldots,a_2^{k_2}\}$, $\ldots$, $C_n=\{a_n^1,a_n^2,\ldots,a_n^{k_n}\}$ form a partition of A' and the equivalence defined here is the P-equivalence defined earlier. Hence the result follows by a previous proposition saying that the number of different P-equivalence classes of A_N is

$$\binom{N}{k_1,k_2,\ldots,k_{n-1}}=\frac{N!}{k_1!\,k_2!\cdots k_n!},\quad N=k_1+\cdots+k_n.$$

■

Corollary Suppose that $x_1,x_2,\ldots,x_n$ are real numbers and N is a positive integer. Then

a) $$(x_1+\cdots+x_n)^N=\sum_{k_1=0}^{N}\sum_{k_2=0}^{N_1-k_1}\cdots\sum_{k_{n-1}=0}^{N-k_1-\cdots-k_{n-2}}\frac{N!}{k_1!\,k_2!\cdots k_n!}\,x_1^{k_1}x_2^{k_2}\cdots x_n^{k_n},$$

where $k_1+k_2+\cdots+k_n=N$. In particular,

b) $$(x+y)^N=\sum_{k=0}^{N}\binom{N}{k}x^ky^{N-k}$$

Proof: Consider the equation

(c) $$(x_{1,1}+x_{1,2}+\cdots+x_{1,n})(x_{2,1}+x_{2,2}+\cdots+x_{2,n})\cdots(x_{N,1}+x_{N,2}+\cdots+x_{N,n})$$

$$=\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\cdots\sum_{l_N=1}^{n}x_{1,l_1}\cdot x_{2,l_2}\cdots x_{N,l_N},$$ where all x_i^j

are real numbers. If we put $x_{1,1}=x_{2,1}=\cdots=x_{N,1}=x_1$

$x_{1,2}=x_{2,2}=\cdots=x_{N,2}=x_2$

$\vdots$

$x_{1,n}=x_{2,n}=\cdots=x_{N,n}=x_n$

then (c) reduces to the L.H.S. of (a). In this case $x_{1,\ell_1} x_{1,\ell_2} \cdots x_{N,\ell_N} = x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}$ if and only if in $x_{1,\ell_1}, x_{1,\ell_2}, \dots, x_{N,\ell_N}$ considered as a word the letter x_1 occurs k_1 times, x_2 occurs k_2 times, etc. The number of terms at the right hand side of (c) which are equal to $x_1^{k_1} \cdots x_n^{k_n}$ is then equal to the number of words of length N, in which $x_1, \dots, x_n$ appear $k_1, k_2, \dots, k_n$ times, respectively. ∎

<u>Remark</u>: As immediate consequence we have

i) $\sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{k} 1^k \cdot 1^{n-k} = (1+1)^n = 2^n$.

ii) $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = \sum_{k=0}^{n} (-1)^k 1^{n-k} \binom{n}{k} = (1-1)^n = 0$.

iii) $\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$. To see this note that any subset of k-elements of the set $\{x_1, \dots, x_n, x_{n+1}\}$ either contains x_{n+1} or does not contain x_{n+1}. In former case there $\binom{n}{k-1}$ subsets (i.e., $k-1$ element subsets of $\{x_1, \dots, x_n\}$) and in the latter case there are $\binom{n}{k}$ subsets (i.e., k element subsets of $\{x_1, \dots, x_n\}$).

iv) $\binom{n+m}{k} = \sum_{j=0}^{\min(n,k)} \binom{n}{j}\binom{m}{k-j}$.

<u>Proof</u> Just consider choosing a team of k-students from a class of n boys and m girls. ∎

v) $\binom{-n}{k} \doteq \frac{-n(-n-1)\cdots(-n-k+1)}{k!} = (-1)^k \binom{n+k-1}{k}$

vi) $\sum_{\substack{k_1+\cdots+k_\ell=n \\ k_i \geq 0}} \frac{n!}{k_1! k_2! \cdots k_\ell!} 1^{k_1} 1^{k_2} \cdots 1^{k_\ell} = (\underbrace{1+1+\cdots+1}_{\ell\text{-times}})^n = \ell^n$.

Video 19

Proposition: Let $S=\{s_1,s_2,\dots,s_n\}$ be a set of positive integers and consider the Cartesian product

$$S^N=\{s_1,s_2,\dots,s_n\}\times\{s_1,s_2,\dots,s_n\}\times\cdots\times\{s_1,s_2,\dots,s_n\}$$

Let $E_k=\{(s_{i_1},s_{i_2},\dots,s_{i_N})\in S^N \mid s_{i_1}+\dots+s_{i_N}=k\}$.

Then the number of elements in E_k equals the coefficient of t^k in
$$(t^{s_1}+t^{s_2}+\dots+t^{s_n})^N = \sum_{k=0}^{\infty} a_k t^k$$

Proof: Proof is similar to that of the previous Corollary. Just put $x_1=t^{s_1}, x_2=t^{s_2},\dots, x_n=t^{s_n}$. Then in (c) $x_{1,\ell_1}x_{1,\ell_2}\cdots x_{N,\ell_N}=t^k$ if and only if $k_1s_1+k_2s_2+\dots+k_ns_n=k$. The number of such terms in (c) is a_k. ∎

Example: In the above proposition let $S=\{1,2,\dots,6\}$. So by the proposition the number of ways of getting the sum equal to k with numbers shown by n dice is the coefficient of t^k in $(t+t^2+\dots+t^6)^n$. For $|t|<1$, the polynomial is equal

$$(t+\dots+t^6)^n = t^n(1+t+\dots+t^5)^n \qquad 1+t+\dots+t^5=\frac{1-t^6}{1-t}$$
$$= t^n(1-t^6)^n(1-t)^{-n}$$
$$= t^n\sum_{i=0}^{6}(-1)^i\binom{n}{i}t^{6i}\sum_{j=0}^{\infty}(-1)^j\binom{-n}{j}t^j \quad ?$$
$$= t^n\sum_{i=0}^{6}(-1)^i\binom{n}{i}t^{6i}\sum_{j=0}^{\infty}\binom{n+j-1}{n-1}t^j$$
$$= \sum_{i=0}^{6}(-1)^i\binom{n}{i}\sum_{j=0}^{\infty}\binom{n+j-1}{n-1}t^{j+6i+n}$$

Let $k=n+6i+j$ then $n+j=k-6i$. So the coefficient of t^k is the

$$\sum_{0 \le i \le \frac{k-1}{6}} (-1)^i \binom{n}{i} \binom{k-6i-1}{n-1}.$$

Definition: We'll say that two words are permutation equivalent if they contain exactly the same letters counted with multiplicity.

Proposition: Let A be an alphabet having n letters and let $\{C_1, \dots, C_i\}$ be a partition of A, where C_j contains n_j elements, $j = 1, \dots, i$. Let M be the set of all words in H, containing $k_1, k_2, \dots, k_i$ letters from $C_1, \dots, C_i$, respectively. Then the number of different permutation equivalence classes equals
$$\binom{n_1}{k_1}\binom{n_2}{k_2} \cdots \binom{n_i}{k_i}.$$

Proof: Easy and left as an exercise.

Proposition: Let $A = \{a_1, \dots, a_n\}$ be an alphabet and let $W \subseteq H_2$ be the class of all words of length k. Then the number of all permutation equivalence classes in W equals $\binom{n+k-1}{n-1}$.

Proof: Let O be the set of all words in W, whose k_1 first letters coincide with a_1, k_2 succeeding letters with $a_2, \dots,$ and finally k_n last letters with a_n, for all choices $k_1, \dots, k_n$ s.t. $k_i \geq 0$ and $k_1 + \dots + k_n = k$. Note that the equivalence class of any such word has a unique element of the form

$$\underbrace{a_1 \cdots a_1}_{k_1} \underbrace{a_2 \cdots a_2}_{k_2} \cdots \underbrace{a_n \cdots a_n}_{k_n}.$$ Hence, we just need to count such words.

Video 20

Moreover, the set of such words is in one to one correspondence with the set of all sequences of the form
$\underbrace{0\cdots0}_{k_1}1\underbrace{0\cdots0}_{k_2}1\cdots\cdots1\underbrace{0\cdots0}_{k_n}$. Finally, the number of such sequences is $\binom{n+k-1}{n-1}$ since any such word is obtained by replacing $n-1$ zeros by 1 in the word $0\cdots0$ consisting of $n+k-1$ zeros. This finishes the proof. ∎

Definition: A collection of different objects is called a population and the number of elements of a population is called its size. Choosing objects from a population is called a sampling. A sequence of chosen elements is a sample and the number of elements forming a sample is called its size.

Note that any finite population can be regarded as an alphabet. In this case, a sample of size k is a word of length k.

Definition: If in a sampling the chosen objects are not put back into the population, then the sampling is said to be without replacement, and if they are put back then the sampling is said to be with replacement.

Consequently, in case every object is considered as a letter, the sampling without replacement of a finite population corresponds to forming the words in H_1, and the sampling with replacement corresponds to forming the words of H_2.

We'll consider four types of samplings:

A sampling of type S_1 will be a sampling without replacement in which two samples with the same size are different if and only if they contain a certain object in different places, i.e., the order is important (Example: $abce \neq abcd \neq adcb$)

A sampling of type S_2 will be a sampling without replacement in which two samples are considered to be different if and only if they contain different sets of objects, i.e., the order is not important.
(Example: $abcd = adcb$ but $abcd \neq abcf$)

A sampling of type S_3 will be a sampling with replacement in which two samples are different if and only if they contain different objects at a certain point, i.e., the order is important.
(Example: $aabc \neq abac$)

A sampling of type S_4 will be a sampling with replacement in which two samples are different if and only if they contain a certain object in different quantities, i.e., the order is not important. (Example: $aabc \neq aabb$ but $aabc = abca$)

Next we'll interpret the results of the preceding propositions using the above terminology.

(I) $n!$

In a sampling of type S_1, when the population is of size n, the number of samples of size n equals (I)

(II) $n^{(k)} = \dfrac{n!}{(n-k)!}$ $(k \leq n)$

In a sampling of type S_1, where the population size is n, the number of samples of size k equals (II).

(II) $\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!}$ $\quad (k \leq n)$

In a sampling of type S_2, where the population is of size n, the number of samples of size k equals (III).

Example a) A club with 20 members has to appoint a four people direction committee. The number of possible committees is $\binom{20}{4}$.

b) The number of different poker hands "5 cards from an ordinary deck of 52 cards" is $\binom{52}{5}$.

(IV) $\dbinom{n}{k_1, k_2, \dots, k_{r-1}} = \dfrac{n!}{k_1!\,k_2!\,\cdots\,k_r!}$, when $k_1 + \dots + k_r = n$.

In a sampling of type S_3, where the population is of size r, the number of different samples of size n in which the first, the second, ..., and the last object figure $k_1, k_2, \dots, k_r$ times, respectively, equals (IV).

Example: The number of ways of placing 4 red, 2 white and 2 black balls into 8 boxes is

$8! \,/\, 4!\,2!\,2!$

(V) $\binom{n_1}{k_1}\binom{n_2}{k_2}\cdots\binom{n_r}{k_r}$

Video 21

In a sampling of type S_2 of a population containing $n_1, n_2, \ldots, n_i$ different objects belonging to different categories $C_1, C_2, \ldots, C_i$ respectively, where $\{C_1, \ldots, C_i\}$ is a partition of a population, the number of different samples in which there are $k_1, k_2, \ldots, k_i$ objects from $C_1, C_2, \ldots, C_i$ respectively, equals (V).

(VI) n^k

In a sampling of type S_3, when the population is of size n, the number of samples of size k equals (VI).

Example: The number of different numbers that can be written with 5 digits is 10^5.

(VII) $\binom{k+n-1}{n-1}$

In a sampling of type S_4, when the population is size n, the number of samples of size k equals (VII).

Example: The number of partial derivatives of order k of an analytic function $f(x_1, \ldots, x_n)$ of n variables is equal to VII. This number is also equal to the number of different monomials of degree k in n variables $x_1, \ldots, x_n$.

§2.2. Equally Likely Outcomes:

Multinomial Distribution: Assume that Ω is a union of mutually exclusive events $\Omega = E_1 \cup E_2 \cup \cdots \cup E_i$, and $\mathbb{P}(E_i) = p_i$, $i = 1, \ldots, n$ $\left(\sum_i p_i = 1\right)$. Suppose that n independent identical experiments are carried out at the outcome of which the same events are likely to occur. What is the probability that, during these independent experiments $E_1, \dots, E_i$ occur $k_1, \dots, k_i$ times, respectively with $k_1 + \dots + k_n = n$?

<u>Answer</u>, Suppose that our experiment consists in typing a key of typewriter capable of generating i different letters $a_1, \dots, a_i$, at a time. We'll say that $E_1, \dots, E_i$ occurs if the machine types $a_1, a_2, \dots, a_i$, respectively. So the occurrence of a sequence of n independent events corresponds to the occurrence of a word $A = a_{\bar{v}_1} a_{\bar{v}_2} \dots a_{\bar{v}_n}$ of independent letters from the alphabet $a_1, a_2, \dots, a_i$. Then

$\mathbb{P}(A) = p(a_{\bar{v}_1}) p(a_{\bar{v}_2}) \dots p(a_{\bar{v}_n})$. If $a_1, a_2, \dots, a_i$ appear $k_1, k_2, \dots, k_i$, times in A, respectively, then $\mathbb{P}(A) = p_1^{k_1} p_2^{k_2} \dots p_i^{k_i}$. The number of such words is equal to $n! / k_1! k_2! \dots k_i!$. Thus, the probability of the set of words in which the letters $a_1, a_2, \dots, a_i$ appear $k_1, k_2, \dots, k_i$, times respectively, equal

$$\frac{n!}{k_1! k_2! \dots k_i!} p_1^{k_1} \dots p_i^{k_i}. \qquad (p_j \geq 0,\ p_1 + \dots + p_i = 1)$$

Here $k_1, \dots, k_i$ are non negative integers with $k_1 + \dots + k_i = n$. Let $\Omega = \{(k_1, \dots, k_i) \mid k_j \geq 0\ \forall j,\ k_1 + \dots + k_n = n\}$.

Define the probability measure p on Ω as

$$p(k_1, \dots, k_i) = \frac{n!}{k_1! \dots k_i!} p_1^{k_1} \dots p_i^{k_i}.$$

Then $$\sum_{(k_1, \dots, k_n) \in \Omega} p(k_1, \dots, k_i) = 1.$$

Example 1) A fair die is thrown 10 times. The probability that 1 occurs 2 times, 2 occurs 2 times, 3 occurs 3 times and 5 occurs 3 times equals

$$\frac{10!}{2!\,2!\,3!\,3!}\left(\frac{1}{6}\right)^{10}.$$

2) 10 different balls are placed "at random" into three boxes A, B and C. What is the probability that there are 3 balls in one box, 3 in another box and 4 in the third box?

Answer: The probability that there are 3 balls in A, 3 balls in B and 4 balls in C equals

$$\frac{10!}{3!\,3!\,4!}\left(\frac{1}{3}\right)^{10}. \qquad \underset{A\ \ B\ \ C}{(3,3,4)}$$

However the situation (3,4,3) and (4,3,3) also give the solution. Hence the result is

$$3 \times \frac{10!}{3!\,3!\,4!}\left(\frac{1}{3}\right)^{10}.$$

Example: What is the probability that among 10 people 3 people have the same birthday, 2 others have another same birthday, and 5 others have different birthdays. (7 different days are to be distributed among 10 people). (Not to born on a leap year!)

Answer: We have $m = 365$, $\alpha_0 = 358$, $\alpha_1 = 5$, $\alpha_2 = 1$, $\alpha_3 = 1$. The required probability is

$$\frac{365!}{357!\;5!\;1!\;1!}\;\frac{10!}{3!\;2!\;1!\,1!\,1!\,1!\,1!}\cdot\frac{1}{(365)^{10}}$$

Video 22

Binomial Distribution:

Suppose that independent and identical experiments are carried out. If the probability of an event E is p, then the probability of k success (observing E k times) over n independent experiments equal

$\rightarrow$ $p_k(n,p) = \binom{n}{k} p^k q^{n-k}$, where $p+q=1$

Let $\Omega = \{0,1,2,\dots,n\}$ and $\mathcal{A}$ be the Boolean algebra of all subsets of Ω. Then $\mathbb{P}(\{k\}) \doteq p_k(n,p)$ is sufficient to define a probability for every event in $\mathcal{A}$. $p_k(n,p)$ defined on Ω is called the binomial distribution with parameters (n,p).

Note that

$$\sum_{k=0}^{n} \mathbb{P}(\{k\}) = \sum_{k=0}^{n} \binom{n}{p} p^k q^{n-k} = (p+q)^n = 1.$$

$$\mathbb{P}(\{k,l\}) = \mathbb{P}(\{k\}) + \mathbb{P}(\{l\}).$$

Examples: 1) A coin is tossed n times independently. What is the probability of having k heads when the probability of a coin to turn head is p?

Let E be the event $E =$ "head occurs" and $p = \mathbb{P}(E)$. Then the required probability is $P_k(n,p)$.

2) A die is thrown 10 times, what is the probability that 6 occurs at least once?

Let $E = \{6 \text{ occurs}\}$, with $\mathbb{P}(E) = 1/6$. So the required probability is

$$\sum_{k=1}^{10} \binom{10}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k} = 1 - \left(\frac{5}{6}\right)^{10}$$

Negative Binomial Distribution:

An infinite sequence of independent identical experiments is carried out at the outcome of which the occurrence of an event E, with probability p is observed. What is the probability of the n^{th} success at the $(n+k)^{\text{th}}$ trial (experiment)?

Note that this probability is equal to the probability of $(n-1)$-successes during $(n+k-1)$ first trials $\times$ the probability of a success at the $(n+k)^{\text{th}}$ trial.

$$\binom{n+k-1}{n-1} p^{n-1} q^k \cdot p = \binom{n+k-1}{n-1} p^n q^k.$$

Again note that

$$\sum_{k=0}^{\infty} \binom{n+k-1}{n-1} p^n q^k = p^n \sum_{k=0}^{\infty} \binom{n+k-1}{n-1} q^k$$

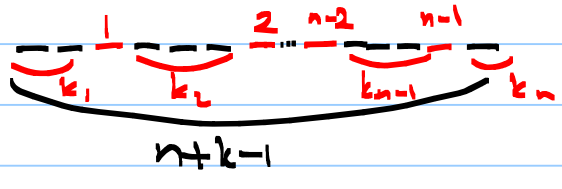
$$= p^n (1-q)^{-n}$$

$$= p^n p^{-n} = 1.$$

Here we use the following that

$$(1-q)^{-n} = \frac{1}{(1-q)^n} = (1+q+q^2+\cdots)^n = \sum_{k=0}^{\infty} \binom{n+k-1}{n-1} q^k,$$

because the number of all possible choices of $k_1, k_2, \ldots, k_n$ such that $k_i \geq 0 \ \forall i$, and $k_1+k_2+\cdots+k_n = k$ is equal $\binom{n+k-1}{n-1}$.

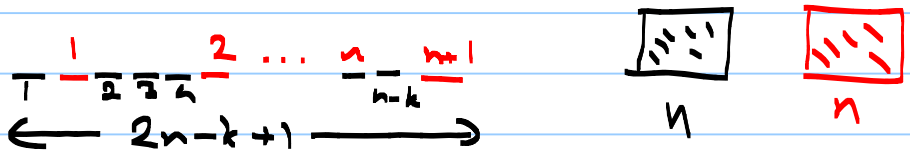


Video 23

Example: (Banach's match-box problem)

A certain mathematician always carries one match box in his right pocket and another in his left. When he wants a match, he selects a pocket at random (with probability $1/2$). Suppose each box initially contained n matches and consider the moment, when the first time, our mathematician discovers that a box is empty. At that moment the other box may contain $0, 1, 2, \ldots, n$ matches. What is the probability that the other box contain k matches?

Answer: $\binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k}$.



Number of choices of n places out of $2n-k$ places is $\binom{2n-k}{n}$. Moreover, each choice occurs with probability $(1/2)^{2n-k+1}$. Black-Red symmetry implies that the answer is

$$2 \cdot \binom{2n-k}{n} \cdot \left(\frac{1}{2}\right)^{2n-k+1} = \binom{2n-k}{n}\left(\frac{1}{2}\right)^{2n-k}.$$

Hypergeometric Distributions

Consider a population containing $n_1, n_2, \ldots, n_r$ different objects belonging to different categories $e_1, e_2, \ldots, e_r$, respectively, where $\{e_1, e_2, \ldots, e_r\}$ is a partition of the population.

In a sampling of type S_2 the probability that a sample of size k containing $k_1, k_2, \ldots, k_r$ objects $(k_1 + k_2 + \cdots + k_r = k)$ belonging to $e_1, e_2, \ldots, e_r$, respectively, equals

$$\frac{\binom{n_1}{k_1}\binom{n_2}{k_2}\cdots\binom{n_i}{k_i}}{\binom{n_1+n_2+\cdots+n_i}{k_1+k_2+\cdots+k_i}}.$$

In particular let $\{e_1, e_2\}$ be a partition of a population of size n, with $n_1+n_2=n$. The probability that a sample of size k containing k_1 objects from e_1 and k_2 objects from e_2, with $k_1+k_2=k$, is equal then

$$\frac{\binom{n_1}{k_1}\binom{n_2}{k_2}}{\binom{n_1+n_2}{k_1+k_2}} = \frac{\binom{n_1}{k_1}\binom{n-n_1}{k-k_1}}{\binom{n}{k}}.$$

Let $\Omega = \{0, 1, \ldots, n\}$. Then the above probability defines a probability on the Boolean algebra of all subsets of Ω, called a hypergeometric distribution:

If $A \subseteq \Omega$, with $|A| = k_1$ then we define

$$\mathbb{P}(A) = \binom{n_1}{k_1}\binom{n-n_1}{k-k_1} \Big/ \binom{n}{k}.$$

(Here n, n_1 and k are fixed.)

Example: 1) An urn contains 10 white balls and 6 black balls. Four balls are drawn without replacement. What is the probability that at least one white ball is obtained?

Answer: $\displaystyle\sum_{k=1}^{4} \frac{\binom{10}{k}\binom{6}{4-k}}{\binom{16}{4}}.$

2) A lot contains n articles. It is known that r of the articles are defective. What is the probabi-

lity that the k^{th} article drawn will be the last defective one in a sampling without replacement? $(k \geq r)$

$\underline{\quad}\ \underline{1}\ \underline{\quad}\ \underline{2}\ \dots\ \underline{\quad}\ \underline{r}$
$\vdash$ k choices $\dashv$

Red ones are defective. We know that the last one is also defective. So then the number of choices is
$\binom{k-1}{r-1}$. So the probability is equal

$$\frac{\binom{k-1}{r-1}}{\binom{n}{r}} = \frac{\text{\# of choices of } r-1 \text{ elements from a set of } k-1 \text{ elements}}{\text{\# of choices of } r \text{ elements from a set of } n \text{ elements}}$$

This number is equal

$$\frac{\binom{r}{r-1}\binom{n-r}{k-r}}{\binom{n}{k-1}} \cdot \frac{1}{n-k+1}, \text{ because}$$

$\binom{r}{r-1}\binom{n-r}{k-r} / \binom{n}{k-1}$ is the probability that a sample of size $k-1$ contains $r-1$ defective objects from a population of size n, containing r defective ones and $1/_{n-k+1}$ is the probability that the last defective one is chosen from the remaining $n-k+1$ objects.

Sample without repetition in a sampling with replacement.

Consider a population of n objects. Suppose that a sample of size k is drawn with replacement. Then the probability in this sample no object is repeated is equal to

$$\frac{n^{(k)}}{n^k} = \frac{n!}{(n-k)!} \left(\frac{1}{n}\right)^k.$$

Example: Suppose that there k people in a room. What is the probability that no two people have the same birthday?

Answer:
$$\frac{365^{(k)}}{365^k} = \frac{365 \cdot 364 \cdot \dots \cdot (365-k+1)}{365 \cdot 365 \cdot \quad \cdot 365}$$

$$= 1 \cdot \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{k-1}{365}\right)$$

Note that if $k \geq 366 > 365$ the probability is zero. If $k = 23$ then the probability is equal to

$1 - \frac{506}{730} < 1/2$. Hence, the probability that two people have the same birthday is greater than $1/2$ if $k \geq 23$.

Video 24

Probability of Drawing Every Object of a Population In a Sampling with Replacement:

Let $\{a_1, \ldots, a_n\}$ be a given population. Suppose a sample of size m ($m \geq n$) drawn with replacement. The probability that every object $a_1, \ldots, a_n$ is drawn at least once is equal to the probability that none of objects is missing. For $i = 1, \ldots, n$, let E_i be the event that a_i is missing. Then

$$\mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_k}) = \left(\frac{n-k}{n}\right)^m, \quad i_1 < i_2 < \cdots < i_k \leq n.$$

Hence,
$$\mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < \cdots < i_k} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_k})$$
$$= \sum_{k=1}^{n} (-1)^{k+1} \left(\frac{n-k}{n}\right)^m \binom{n}{k}$$

Hence, the probability that none of the objects is missing is
$$1 - \mathbb{P}\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(\frac{n-k}{n}\right)^m$$

Complements: i) Let $\{E_1, \ldots, E_n\}$ be a finite family of events and let $\mathbb{P}_{[k]}$ be the probability that exactly k of the n events $E_1, \ldots, E_n$ occur. Then

$$\mathbb{P}_{[k]} = \binom{k}{k} \sum_{i_1 < \cdots < i_k} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_k})$$
$$- \binom{k+1}{k} \sum_{i_1 < \cdots < i_{k+1}} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_{k+1}})$$
$$+ \binom{k+2}{k} \sum_{i_1 < \cdots < i_{k+2}} \mathbb{P}(E_{i_1} \cap \cdots \cap E_{i_{k+2}})$$
$$\vdots$$
$$\pm \binom{n}{k} \mathbb{P}(E_1 \cap \cdots \cap E_n).$$

ii) Similarly, if P_k denotes the probability that at least k of the events $E_1, E_2, \ldots, E_n$ occur, then

$$P_k = \binom{k-1}{k-1} \sum_{i_1<\cdots<i_k} \mathbb{P}(E_{i_1}\cap\cdots\cap E_{i_k})$$

$$- \binom{k}{k-1} \sum_{i_1<\cdots<i_{k+1}} \mathbb{P}(E_{i_1}\cap\cdots\cap E_{i_{k+1}})$$

$$+ \binom{k+1}{k-1} \sum_{i_1<\cdots<i_{k+2}} \mathbb{P}(E_{i_1}\cap\cdots\cap E_{i_{k+2}})$$

$$\vdots$$

$$\pm \binom{n-1}{k-1} \mathbb{P}(E_1\cap\cdots\cap E_n).$$

If $k=1$, then we obtain

$$P_1 = \sum_{j=1}^{n} (-1)^{j+1} \sum_{i_1<\cdots<i_j} \mathbb{P}(E_{i_1}\cap\cdots\cap E_{i_j}).$$

iii-a) Suppose that the events $E_1, \ldots, E_n$ are such that the probability of the occurrence of any specified k of them is p_k, $k=1,\ldots,n$. Then by the above formula, the probability of the occurrence of at least one of them is

$$P_1 = \binom{n}{1} p_1 - \binom{n}{2} p_2 + \cdots + (-1)^{n+1} p_n$$

$$\mathbb{P}(E_i) \qquad \mathbb{P}(E_i\cap E_j)$$

iii-b) Suppose that an urn contains n objects $a_1, \ldots, a_n$. The objects are drawn with replacement, we say that there is a match if a_j is the j^{th} drawn object. Then the probability that there is at least one match in a sample of size n is

$$\sum_{k=1}^{n} (-1)^{k+1} 1/k! \qquad \text{(Exercise).}$$

Video 25

CHAPTER 3: Random Variables and their Distributions:

§3.1. More about σ-algebras and Probability:

We start by asking if it is possible to construct independent class of sub-σ-algebras.

Definition: Let $(\Omega', \mathcal{A}')$ and $(\Omega'', \mathcal{A}'')$ be two measurable spaces and let $\Omega = \Omega' \times \Omega''$. Then the subsets of Ω of the form $A' \times A''$ with $A' \in \mathcal{A}'$ and $A'' \in \mathcal{A}''$ are called measurable rectangles. The collection of all rectangles is not a σ-algebra. The algebra of subsets of Ω generated by all measurable rectangles is called the product σ-algebra of $\mathcal{A}'$ and $\mathcal{A}''$ and is denoted by $\mathcal{A}' \otimes \mathcal{A}''$, in other words $\mathcal{A}' \otimes \mathcal{A}''$ is generated by $\{A' \times A'' \mid A' \in \mathcal{A}', A'' \in \mathcal{A}''\}$.

Proposition: If $\mathcal{B}$ denotes the Borel σ-algebra on $\mathbb{R}$ then $\mathcal{B}^2 = \mathcal{B} \otimes \mathcal{B}$, i.e., the Borel σ-algebra of $\mathbb{R}^2$ is the product of σ-algebras.

Proof. $\mathcal{B} \otimes \mathcal{B}$ is the σ-algebra generated by the set of all rectangles, i.e., the sets of the following type: $A' \times A''$, where $A', A'' \in \mathcal{B}$. On the other hand, $\mathcal{B}^2$ is the σ-algebra generated by the family of all open subsets of $\mathbb{R}^2$. Indeed, if $S' = \{(-\infty, x_1] \times (-\infty, x_2] \mid x_1, x_2 \in \mathbb{Q}\}$ then $\mathcal{B}^2 = \sigma(S') = \sigma(S \times S)$, where $S = \{(-\infty, x] \mid x \in \mathbb{Q}\}$.
Clearly, $S' \subseteq \mathcal{B} \otimes \mathcal{B}$ and hence $\mathcal{B}^2 = \sigma(S') \subseteq \mathcal{B} \otimes \mathcal{B}$.

On the other hand, $\mathcal{B} \otimes \mathcal{B}$ is generated by measurable rectangles of the form $A' \times A''$, where $A' \in \mathcal{B}$ and $A'' \in \mathcal{B}$. So, A' and A'' are countable unions of elements of S. Hence $A' \times A''$ is a union of elements of S' and hence $A' \times A'' \in \mathcal{B}^2$. This implies that $\mathcal{B} \otimes \mathcal{B} \subseteq \mathcal{B}^2$. Hence, $\mathcal{B}^2 = \mathcal{B} \otimes \mathcal{B}$. ∎

Proposition: Let $(\Omega', \mathcal{A}', \mathbb{P}')$ and $(\Omega'', \mathcal{A}'', \mathbb{P}'')$ be two probability spaces. Then there is a unique probability measure on $\mathcal{A}' \otimes \mathcal{A}''$, denoted by $\mathbb{P} = \mathbb{P}' \times \mathbb{P}''$ such that

$$\mathbb{P}(A' \times A'') = (\mathbb{P}' \times \mathbb{P}'')(A' \times A'') = \mathbb{P}'(A')\, \mathbb{P}''(A'').$$

Proof: $\mathbb{P}(\phi) = \mathbb{P}(\phi \times \phi) = \mathbb{P}'(\phi) \times \mathbb{P}''(\phi) = 0.$

$$\mathbb{P}(\Omega) = \mathbb{P}' \times \mathbb{P}''(\Omega' \times \Omega'') = \mathbb{P}'(\Omega')\, \mathbb{P}''(\Omega'') = 1 \cdot 1 = 1.$$

Uniqueness: Let $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ be two probability measures in $\mathcal{A}' \otimes \mathcal{A}''$ so that

$$\mathbb{P}(A' \times A'') = \widetilde{\mathbb{P}}(A' \times A''), \text{ for all } A' \in \mathcal{A}', A'' \in \mathcal{A}''.$$

If $A \in \mathcal{A}' \otimes \mathcal{A}''$ then $A = \bigcup_k A'_k \times A''_k$, $A'_k \in \mathcal{A}'$ and $A''_k \in \mathcal{A}''$. Let $B_n = \bigcup_{k=1}^{n} A'_k \times A''_k$.

Then $B_n \subseteq B_{n+1}$, $\forall n$ and $\bigcup B_n = \bigcup A'_k \times A''_k = A$.

Then $\mathbb{P}(A) = \mathbb{P}(\bigcup B_n) = \lim \mathbb{P}(B_n)$ and

$$\widetilde{\mathbb{P}}(A) = \widetilde{\mathbb{P}}(\bigcup B_n) = \lim \widetilde{\mathbb{P}}(B_n).$$

However $\mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{k=1}^{n} A'_k \times A''_k\right) = \widetilde{\mathbb{P}}\left(\bigcup_{k=1}^{n} (A'_k \times A''_k)\right)$

$$= \widetilde{\mathbb{P}}(B_n)$$

The details of $\mathbb{P}\left(\bigcup_{k=1}^{n} A'_k \times A''_k\right) = \widetilde{\mathbb{P}}\left(\bigcup_{k=1}^{n} A'_k \times A''_k\right)$

and the proof of σ-additivity of $\mathbb{P}$ are left as an exercise.

__Hint:__ Any element of the algebra $\mathcal{A}' \otimes \mathcal{A}''$ can be written as the disjoint union countably many measurable rectangles. ∎

Video 26

Definition: Let $(\Omega', \mathcal{A}', \mathbb{P}')$, $(\Omega'', \mathcal{A}'', \mathbb{P}'')$ and $(\Omega, \mathcal{A}, \mathbb{P})$ be as above. Then $(\Omega, \mathcal{A}, \mathbb{P})$ is called the product probability space and $\mathbb{P} = \mathbb{P}' \times \mathbb{P}''$ the product probability measure on $\mathcal{A}' \otimes \mathcal{A}'' = \mathcal{A}$.

Example: Let $(\Omega', \mathcal{A}', \mathbb{P}')$ and $(\Omega'', \mathcal{A}'', \mathbb{P}'')$ denote the probability spaces which represents tossing a coin and rolling a die, respectively. So, $\Omega' = \{h, t\}$ and $\Omega'' = \{1, 2, 3, 4, 5, 6\}$. $\mathcal{A}' = \mathcal{P}(\Omega')$, $\mathcal{A}'' = \mathcal{P}(\Omega'')$. $\Omega = \Omega' \times \Omega''$ and

$\mathbb{P}(\{(x,y)\}) = \mathbb{P}'(\{x\})\, \mathbb{P}''(\{y\}) = 1/2 \cdot 1/6 = 1/12$, for any $(x,y) \in \Omega$ (assuming the classical scheme for both Ω' and Ω''.)

Theorem: Let $(\Omega', \mathcal{A}')$ and $(\Omega'', \mathcal{A}'')$ be the measurable spaces and let $\mathcal{A}$ denote product algebra $\mathcal{A}' \otimes \mathcal{A}''$. Define, $\mathcal{A}^{(1)} = \{A' \times \Omega'' \mid A' \in \mathcal{A}'\}$ and
$\mathcal{A}^{(2)} = \{\Omega' \times A'' \mid A'' \in \mathcal{A}''\}$. Then $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ are two independent sub-σ-algebras of $\mathcal{A}$.

Proof: It is easy to see that $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ are sub-algebras:

1) $\emptyset = \emptyset \times \Omega'' \in \mathcal{A}^{(1)}$ and $\emptyset = \Omega' \times \emptyset \in \mathcal{A}^{(2)}$.
2) $\Omega = \Omega' \times \Omega'' \in \mathcal{A}^{(1)}$ and $\Omega = \Omega' \times \Omega'' \in \mathcal{A}^{(2)}$.
3) Let $B_k = A'_k \times \Omega''$, $k \in \mathbb{N}$. Then $B_k \in \mathcal{A}^{(1)}$ and
$$\bigcup_{k\in\mathbb{N}} B_k = \bigcup_{k\in\mathbb{N}} A'_k \times \Omega'' = \underbrace{\Big(\bigcup_{k\in\mathbb{N}} A'_k\Big)}_{\in \mathcal{A}'} \times \Omega'' \in \mathcal{A}^{(1)}.$$

A similar argument works for $\mathcal{A}^{(2)}$.

Next, we need to show that $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ are independent: Take any $A^{(1)} = A' \times \Omega''$ and $A^{(2)} = \Omega' \times A''$, for some $A' \in \mathcal{A}'$ and $A'' \in \mathcal{A}''$. Then
$A^{(1)} \cap A^{(2)} = (A' \times \Omega'') \cap (\Omega' \times A'') = A' \times A''$.

Hence, $\mathbb{P}(A^{(1)} \cap A^{(2)}) = \mathbb{P}(A' \times A'')$
$$= \mathbb{P}'(A') \cdot \mathbb{P}''(A'')$$
$$= \mathbb{P}(A' \times \Omega'') \; \mathbb{P}(\Omega' \times A'').$$
This finishes the proof. ∎

Proposition: Let $\mathbb{P}$ be a probability measure on $\mathcal{A}' \otimes \mathcal{A}''$. Define the following functions on $\mathcal{A}'$ and $\mathcal{A}''$ as follows:
$\mathbb{P}'(A') \doteq \mathbb{P}(A' \times \Omega''), \; \forall A' \in \mathcal{A}'$, and
$\mathbb{P}''(A'') \doteq \mathbb{P}(\Omega' \times A''), \; \forall A'' \in \mathcal{A}''$.

Then $\mathbb{P}'$ and $\mathbb{P}''$ are probability measures.

Proof: Just prove that $\mathbb{P}'$ is a probability measure on $\mathcal{A}'$:
1) $\mathbb{P}'(\emptyset) = \mathbb{P}(\emptyset \times \Omega'') = \mathbb{P}(\emptyset) = 0.$
2) $\mathbb{P}'(\Omega_1) = \mathbb{P}(\Omega_1 \times \Omega'') = \mathbb{P}(\Omega) = 1.$
3) If $A_k \in \mathcal{A}'$, $k \in \mathbb{N}$ are disjoint subsets then

$$\mathbb{P}'\left(\bigcup_{k=1}^{\infty} A_k\right) = \mathbb{P}\left(\left(\bigcup_{k=1}^{\infty} A_k\right) \times \Omega''\right) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} (A_k \times \Omega'')\right)$$
$$= \sum_{k=1}^{\infty} \mathbb{P}(A_k \times \Omega'')$$
$$= \sum_{k=1}^{\infty} \mathbb{P}'(A_k),$$

where $\bigcup_{k=1}^{\infty}(A_k \times \Omega'')$ is a disjoint union since $\bigcup_{k=1}^{\infty} A_k$ is a disjoint union.

4) $\mathbb{P}'(A) = \mathbb{P}(A \times \Omega'') \geq 0$, for any $A \in \mathcal{A}'$.
∎

Definition: The probabilities $\mathbb{P}'$ and $\mathbb{P}''$ defined in the above proposition are called the marginal probabilities on $\mathcal{A}'$ and $\mathcal{A}''$, respectively.

One can similarly define the product of more than two spaces. Namely, given probability spaces $(\Omega^i, \mathcal{A}^i, \mathbb{P}^i)$, $i=1, \dots, n$, let $(\Omega, \mathcal{A}, \mathbb{P})$ be

$$\left(\prod_{i=1}^{n} \Omega^i, \bigotimes_{i=1}^{n} \mathcal{A}^i, \prod_{i=1}^{n} \mathbb{P}^i\right).$$

Define similarly, $\mathcal{A}^{(i)} = \{\Omega_1 \times \cdots \times A_i \times \cdots \times \Omega_n \mid A_i \in \mathcal{A}_i\}$.

<u>Theorem:</u> $\mathcal{A}^{(i)}$'s are independent σ-algebras of $\mathcal{A}$.

<u>Remark:</u> If $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$, then

$$\mathcal{B}^n = \bigotimes_{i=1}^{n} \mathcal{B}.$$

<u>Remark:</u> (Countable Products)
Let $\{(\Omega_j, \mathcal{A}_j, \mathbb{P}_j)\}_{j=1}^{\infty}$ be a countable family of probability spaces. Then one may define the product probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as follows:

$\Omega = \prod_{j=1}^{\infty} \Omega_j$, $\mathcal{A}$ is the algebra generated by sets of the form $\bigtimes_{j=1}^{\infty} A_j$, $A_j \in \mathcal{A}_j$ and

$\mathbb{P} = \prod_{j=1}^{\infty} \mathbb{P}_j$ given by $\mathbb{P}\left(\bigtimes_{j=1}^{\infty} A_j\right) = \prod_{j=1}^{\infty} \mathbb{P}_j(A_j)$

Before we introduce the notion of Random Variables we need to introduce two concepts.

<u>Definition:</u> Let $(\Omega, \mathcal{A})$ be a measurable space, $\mathcal{S} \subseteq \mathcal{A}$ and $\mathbb{P}_0$ a mapping from $\mathcal{S}$ into $[0,1]$.

Video 27

A probability measure $\mathbb{P}$ on $\mathcal{A}$ is called an extension of $\mathbb{P}_0$ if $\forall s \in \mathcal{S}$, $\mathbb{P}(S) = \mathbb{P}_0(S)$.

Definition: A class $\mathcal{S}$ of subsets of a set Ω is called a Boolean semi algebra if it satisfies the following conditions

a) $\phi, \Omega \in \mathcal{S}$,

b) $\mathcal{S}$ is closed under finite intersections,

c) If $s \in \mathcal{S}$, then S^c is the disjoint union of a finite family of pairwise disjoint subsets of $\mathcal{S}$.

Example: The measurable rectangles of $(\Omega_1, \mathcal{A}_1) \times (\Omega_2, \mathcal{A}_2)$ form a Boolean semi algebra of subsets of $\Omega_1 \times \Omega_2$.

Proof: a) $\phi = \phi \times \phi$ and $\Omega_1 \times \Omega_2$ are measurable rectangles.

b) Intersections of two measurable rectangles is measurable:

$(A_1 \times A_2) \cap (A_1' \times A_2') = (A_1 \cap A_1') \times (A_2 \cap A_2')$.

c) $(A_1 \times A_2)^c = (A_1^c \times A_2) \cup (\Omega_1 \times A_2^c)$, where

$(A_1^c \times A_2) \cap (\Omega_1 \times A_2^c) = \phi$.

Theorem (Extension Theorem)

The Boolean algebra $\mathcal{C}$ generated by a Boolean semialgebra $\mathcal{S}$ of subsets of Ω consists of the unions $\mathcal{C} = \bigcup_{i \in I} S_i$ of finite families $\{S_i / i \in I\}$

of pairwise disjoint subsets of Ω in $\mathcal{S}$. For every additive function $\mathbb{P}_0$ mapping $\mathcal{S}$ into $[0,1]$ such that $\mathbb{P}_0(\Omega) = 1$, the formula

$\mathbb{P}(C) = \sum_{i\in I} \mathbb{P}_1(S_i)$ defines the unique extension $\mathbb{P}_1$ of $\mathbb{P}_0$, which is additive on $\mathcal{C}$.

If the function $\mathbb{P}_0$ is σ-additive on S then $\mathbb{P}_1$ is a probability on $\mathcal{C}$. In this case, there exists a unique probability $\mathbb{P}$ on $\sigma(S)$ which extends $\mathbb{P}_0$.

§3.2. Random Variables:

A random variable (measurable function) is a function of outcome of an experiment with random outcomes having statistical stability.

<u>Definition:</u> Let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be two measurable spaces. A mapping $X: \Omega \to \Omega'$ is said to be measurable if $X^{-1}(A') \in \mathcal{A}$ for all $A' \in \mathcal{A}'$. In this case, we also say that X is a random variable.

<u>Remark:</u> If $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B})$ then a random variable $X: \Omega \to \mathbb{R}$ is called a real valued random variable.

<u>Example</u> Tossing two dice: $\Omega = \{(a,b) \mid 1 \leq a,b \leq 6\}$ and $\mathcal{A} = \mathcal{P}(\Omega)$ and $\mathbb{P}(A) = |A|/|\Omega|$, $\forall A \in \mathcal{A}$, (classical scheme). Let $X: \Omega \to \mathbb{R}$ be the function $X(a,b) = a+b$. Then X is a random variable, any subset of Ω is measurable.

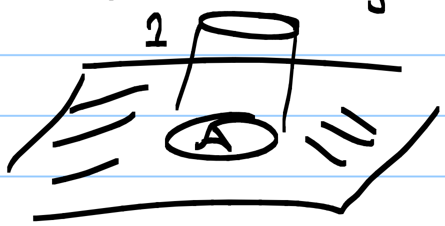
On the other hand, if $\mathcal{A} = \{\emptyset, \Omega\}$ then X would not measurable since $X^{-1}(\{2\}) = \{(1,1)\}$ is not in $\mathcal{A}$.

<u>Some Examples of Measurable Functions</u>

<u>Identity mapping:</u> The identity mapping

$Id: \Omega \to \Omega$ is a measurable function of $(\Omega, \mathcal{A})$.

Indicator functions of Measurable sets Let $(\Omega, \mathcal{A})$ be a measurable space and $A \in \mathcal{A}$ any element.
Define
$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$



In measure theory
$\mathbb{1}_A$ is also called the characteristic function of the measurable set A.

$\mathbb{1}_A : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ is measurable since
$$\mathbb{1}_A^{-1}(B) = \begin{cases} \emptyset & \text{if } \{0,1\} \cap B = \emptyset \\ A & \{0,1\} \cap B = \{1\} \\ A^c & \{0,1\} \cap B = \{0\} \\ \Omega & \{0,1\} \subseteq B \end{cases}$$
Hence, $\mathbb{1}_A$ is a measurable function.

Simple Functions Let $\{A_1, \dots, A_k\}$ be a finite partition of Ω and $a_1, \dots, a_k \in \mathbb{R}$. Then the mapping $X : \Omega \to \mathbb{R}$ given by
$$X(\omega) = \sum_{i=1}^{k} a_i \mathbb{1}_{A_i}(\omega) \text{ is a measurable}$$
function.

Indeed, this is a consequence of a more general result.

i) If X, Y are real random variables then so is $X+Y$

ii) If X is a real random variable then so is cX, for any $c \in \mathbb{R}$.

Proof of i) Enough to consider the measurable sets of the form $(-\infty, a]$.

Video 2P

$$(X+Y)^{-1}((-\infty,a]) = \bigcup_{p\in\mathbb{Q}} A_n, \quad A_n = \{\omega \mid X(\omega)\le p,\ Y(\omega)\le a-p\}.$$

$$A_n = X^{-1}(-\infty,p] \cap Y^{-1}(-\infty,a-p]$$

Exercise: Fill the gaps!

Real Valued Functions on finite Sets:

Let $\Omega = \{\omega_1,\dots,\omega_n\}$ and $\mathcal{A} = \mathcal{P}(\Omega)$. Then any function $f: \Omega \to \mathbb{R}$ is measurable.

Continuous Functions of $\mathbb{R}^m$ into $\mathbb{R}^n$

Let $f: \mathbb{R}^m \to \mathbb{R}^n$ be a continuous function. Then for any open set $A \in \mathbb{R}^n$, $f^{-1}(A)$ is open in $\mathbb{R}^m$. Since the Borel algebra is generated by open subsets $f^{-1}(A)$ will be a Borel set whenever A is.

Projections: Let $\{(\Omega_i, \mathcal{A}_i) \mid i=1,\dots,n\}$ be measurable spaces and consider the projection map

$$\pi_k: \bigotimes_{i=1}^{n} \Omega_i \to \Omega_k, \quad \pi_k(\omega_1,\dots,\omega_n) = \omega_k.$$

Then π_k is measurable, since $\pi_k^{-1}(A) = \Omega_1 \times \dots \times \underset{k\text{th place}}{A} \times \dots \times \Omega_n$ for any $A \in \mathcal{A}_k$.

Proposition: Let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be measurable spaces and $\mathcal{C}'$ be a class of subsets of Ω' generating $\mathcal{A}'$. Then a mapping $X: \Omega \to \Omega'$ is measurable if and only if $X^{-1}(A') \in \mathcal{A}$ for all $A' \in \mathcal{C}'$!

Example Since $\mathcal{B}$ on $\mathbb{R}$ is generated by subsets of the form $[-\infty, x]$, $x \in \mathbb{Q}$, any function $X: \Omega \to \mathbb{R}$

is measurable if and only if $X^{-1}((-\infty, x]) \in \mathcal{A}$, for all $x \in \mathbb{R}$.

<u>Proposition</u>: The sum, product, quotient (when it is defined) of random variables are random variables.

<u>Proof</u>: We consider only the real random variable case. Let X and Y be two real random variables. Then for any $a \in \mathbb{R}$

$$A = (X+Y)^{-1}((-\infty, a]) = \{\omega \in \Omega \mid X(\omega) + Y(\omega) < a\} \in \mathcal{A}$$

because

$$A = \bigcup_{n=1}^{\infty} \{\omega \mid X(\omega) < q_n ,\ Y(\omega) < a - q_n\}, \text{ where}$$

$\mathbb{Q} = \{q_n \mid n = 1, 2, \dots\}$ and each

$$\{\omega \mid X(\omega) < q_n ,\ Y(\omega) < a - q_n\} = \{\omega \mid X(\omega) < q_n\} \cap \{\omega \mid Y(\omega) < a - q_n\}$$
$$= X^{-1}((-\infty, q_n]) \cap Y^{-1}((-\infty, a - q_n]),$$

an intersection of two measurable sets.

Before we consider the product variable XY let's study X^2. Notice that

$$(X^2)^{-1}(-\infty, a) = \phi \text{ if } a < 0, \text{ and}$$

$$(X^2)^{-1}(-\infty, a) = X^{-1}(-\sqrt{a}, \infty) \cap X^{-1}(-\infty, \sqrt{a}).$$

So, $(X^2)^{-1}(A)$ is measurable for any open subset A. It follows that X^2 is measurable.
For the product XY consider the equality

$$XY = \frac{1}{4}\left((X+Y)^2 - (X-Y)^2\right).$$

Since X and Y are measurable so is XY by the above formula.

Let $X(\omega) \neq 0$ for any $\omega \in \Omega$. Then $1/X$ is also a r.v because

$(1/X)^{-1}(-\infty, a) = X^{-1}(1/a, 0)$ if $a < 0$,

$(1/X)^{-1}(-\infty, 0) = X^{-1}(-\infty, 0)$,

$(1/X)^{-1}(0, a) = X^{-1}(1/a, \infty)$, which are all measurable sets.

Proposition: If $X: \Omega \to \Omega'$ and $Y: \Omega' \to \Omega''$ are measurable functions then so is $Y \circ X: \Omega \to \Omega''$.

Proof If $A'' \subseteq \Omega''$ is measurable then

$(Y \circ X)^{-1}(A'') = X^{-1}(Y^{-1}(A''))$ is clearly measurable.

Corollary If $\varphi: \mathbb{R} \to \mathbb{R}$ is a Borel measurable function and $X: \Omega \to \mathbb{R}$ is a real random variable then so is $\varphi \circ X: \Omega \to \mathbb{R}$.

Proposition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(\Omega', \mathcal{A}')$ a measurable space and $X: \Omega \to \Omega'$ a measurable mapping. Define $\mathbb{P}': \mathcal{A}' \to \mathbb{R}$ by $\mathbb{P}'(A') = \mathbb{P}(X^{-1}(A'))$, for all $A' \in \mathcal{A}'$. Then $\mathbb{P}'$ is a probability measure on $\mathcal{A}'$.

Proof: $\mathbb{P}'(A') = \mathbb{P}(X^{-1}(A')) \geq 0$, for all $A' \in \mathcal{A}'$.

$\mathbb{P}'(\phi) = \mathbb{P}(X^{-1}(\phi)) = \mathbb{P}(\phi) = 0$ and

$\mathbb{P}'(\Omega') = \mathbb{P}(X^{-1}(\Omega')) = \mathbb{P}(\Omega) = 1$.

Video 29

Finally, let $A_k' \in \mathcal{A}'$ with $A_k' \cap A_l' = \emptyset$ if $k \neq l$.

Then
$$\mathbb{P}'\left(\bigcup_{k=1}^{\infty} A_k'\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{k=1}^{\infty} A_k'\right)\right)$$
$$= \mathbb{P}\left(\bigcup_{k=1}^{\infty} X^{-1}(A_k')\right)$$
$$= \sum_{k=1}^{\infty} \mathbb{P}(X^{-1}(A_k'))$$
$$= \sum_{k=1}^{\infty} \mathbb{P}'(A_k').$$

§3.3. Distribution Functions of Real Random Variables

Let X be a real random variable (r.r.v.) on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\mathbb{P}_X$ be the probability measure induced by X on the Borel σ-algebra $\mathcal{B}$ of $\mathbb{R}$, defined by,

$$\mathbb{P}_X(B) \doteq \mathbb{P}(X \in B) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in B\}) = \mathbb{P}(X^{-1}(B)).$$

Taking $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B})$ we see by the previous proposition that $\mathbb{P}_X$ is a probability measure on $\mathcal{B}$ of $\mathbb{R}$. $\mathbb{P}_X$ is also called the probability distribution of the random variable $X: \Omega \to \mathbb{R}$.

Example: Let $(\Omega, \mathcal{A}, \mathbb{P})$ represent the classical scheme of tossing a coin. So $\Omega = \{h, t\}$.
Let $X: \Omega \to \mathbb{R}$ be defined as

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = h \\ -1 & \text{if } \omega = t \end{cases}$$

Then
$$\{X \in B\} \doteq \{\omega \in \Omega \mid X(\omega) \in B\} = \begin{cases} \emptyset & \text{if } 1 \notin B, -1 \notin B \\ \{h\} & \text{if } 1 \in B, -1 \notin B \\ \{t\} & \text{if } 1 \notin B, -1 \in B \\ \Omega & \text{if } 1 \in B, -1 \in B. \end{cases}$$

$$\mathbb{P}_X(B)=\mathbb{P}(\{X\in B\})=\begin{cases}0 & \text{if } 1\notin B,\ -1\notin B\\ 1/2 & \text{if } 1\in B,\ -1\notin B\\ 1/2 & \text{if } 1\notin B,\ -1\in B\\ 1 & \text{if } 1\in B,\ -1\in B\end{cases}$$

<u>Definition</u>: Given a r.v. X the function defined by $F:\mathbb{R}\to\mathbb{R}$

$$F(x)=\mathbb{P}_X((-\infty,x])=\mathbb{P}(\{\omega\in\Omega \mid X(\omega)\leq x\})=\mathbb{P}(X^{-1}(-\infty,x]),$$

for any $x\in\mathbb{R}$, is called the probability distribution function of the r.v. X. We'll also denote this function as F_X.

<u>Proposition</u>: Let F be the distribution function of X. Then

1) F is a non-negative and non-decreasing function.
2) F is right continuous.
3) $\lim_{x\to-\infty} F(x)=0$ and $\lim_{x\to\infty} F(x)=1$, which we shall state this as $F(-\infty)=0$ and $F(\infty)=1$.

<u>Proof</u>: 1) $F(x)=\mathbb{P}_X((-\infty,x])\geq 0$, for all $x\in\mathbb{R}$, because $\mathbb{P}_X$ is probability measure.

Now let $x_1\leq x_2$, then

(*) $$F(x_2)-F(x_1)=\mathbb{P}_X((-\infty,x_2])-\mathbb{P}_X((-\infty,x_1])$$
$$=\mathbb{P}_X((x_1,x_2])\geq 0$$ and hence, F is non-decreasing.

2) Let (x_n) be a sequence of real numbers with $x\leq x_{n+1}\leq x_n$, for all n, with $\lim x_n=x$.

We must show $\lim_{n\to\infty} F(x_n) = F(x)$. Now,

$$\lim_{n\to\infty} (F(x_n) - F(x)) = \lim_{n\to\infty} (\mathbb{P}_X((-\infty, x_n]) - \mathbb{P}_X((-\infty, x]))$$

$$= \lim_{n\to\infty} \mathbb{P}_X((x, x_n])$$

$$= \lim_{n\to\infty} \mathbb{P}_X(I_n), \text{ where } I_n = (x, x_n].$$

Notice that $I_{n+1} \subseteq I_n$ and $\lim_n I_n = \bigcap_{n=1}^{\infty} I_n = \bigcap_{n=1}^{\infty} (x, x_n] = \phi$

and thus the above limit equals

$$\mathbb{P}_X(\lim_{n\to\infty} I_n) = \mathbb{P}_X(\phi) = 0$$ (by the Monotone Sequential Continuity Thm)

Hence, $\lim_{n\to\infty} F(x_n) = F(x)$.

3) Let $J_n = (-\infty, x_n]$, where (x_n) is a decreasing sequence with $\lim x_n = -\infty$. We must show

$\lim_{n\to\infty} F(x_n) = 0$. Note that $J_{n+1} \subseteq J_n$, for all n and $\lim J_n = \bigcap_{n=1}^{\infty} J_n = \phi$. Then as above

$$F(-\infty) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}_X(J_n) = \mathbb{P}_X(\lim_n J_n) = \mathbb{P}_X(\phi) = 0.$$

Similarly, let $J_n = (-\infty, x_n]$, where (x_n) is an increasing sequence with $\lim x_n = +\infty$. Then

$$F(\infty) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \mathbb{P}_X(J_n) = \mathbb{P}_X(\lim_n J_n) = \mathbb{P}_X(\mathbb{R}) = 1,$$

since $\lim_n J_n = \bigcup_{n=1}^{\infty} J_n = \mathbb{R}$. ▬

Example (Back to the Tossing a coin example)

Recall that

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = h \\ -1 & \text{if } \omega = t \end{cases}. \quad \text{Then } \{X \le x\} = X^{-1}((-\infty, x]) = \begin{cases} \emptyset & \text{if } x < -1 \\ \{t\} & \text{if } -1 \le x < 1 \\ \Omega & \text{if } x \ge 1. \end{cases}$$

Then,

$$F_X(x) = \begin{cases} 0 & \text{if } x < -1 \\ 1/2 & \text{if } -1 \le x < 1 \\ 1 & \text{if } x \ge 1. \end{cases}$$

Example: Let $(\Omega, \mathcal{A}, \mathbb{P})$ represent rolling a pair of dice. Define the r.v. X as the sum of the numbers shown on the dice, i.e., $X((a,b)) = a+b$, where $\Omega = \{(a,b) \mid 1 \le a, b \le 6\}$.

So,

$$\{X \le x\} = X^{-1}((-\infty, x]) = \begin{cases} \emptyset & \text{if } x < 2 \\ \{(1,1)\} & \text{if } 2 \le x < 3 \\ \{(1,1),(1,2),(2,1)\} & \text{if } 3 \le x < 4 \\ \vdots & \\ \Omega & \text{if } x \ge 12 \end{cases}$$

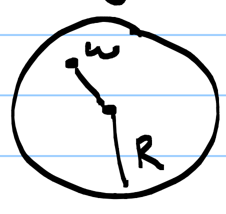
Then

$$F_X(x) = \mathbb{P}(\{X \le x\}) = \begin{cases} 0 & \text{if } x < 2 \\ 1/36 & \text{if } 2 \le x < 3 \\ 1/12 & \text{if } 3 \le x < 4 \\ \vdots & \\ 1 & \text{if } x \ge 12. \end{cases}$$

Example: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the mathematical model of shooting a round target with radius R. Assume that we are in the geometric scheme, i.e., for any event $E \in \mathcal{A}$, $\mathbb{P}(E) = \dfrac{\text{mes } E}{\pi R^2}$

Let X be the r.r.v. that gives the distance to the center.

Ω :



Video 30

So $\{X \le x\} = \begin{cases} \emptyset & \text{if } x < 0 \\ \{\omega = (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \le x^2\} & \text{if } 0 \le x < R \\ \Omega & \text{if } x \ge R. \end{cases}$

Hence $F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x^2/R^2 & \text{if } 0 \le x < R. \\ 1 & \text{if } x \ge R. \end{cases}$

Remark: For a function $F: \mathbb{R} \to \mathbb{R}$ the condition that F is right continuous at x is expressed as $F(x+0) = F(x)$. Similarly, $F(x-0) = F(x)$ will mean that F is left continuous.

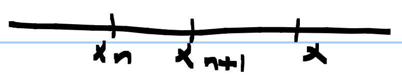
Definition: Let F be the distribution of a r.v. Then a point $x \in \mathbb{R}$ s.t. $F(x-0) = F(x)$ is called a point of continuity of F. If $F(x-0) < F(x)$ then x is called a discontinuity point or a jump point of F. In this case, $F(x) - F(x-0)$ is called the jump of F at x.

Proposition: Let X be a r.v. on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and let F be its distribution function and $\mathbb{P}_X$ the probability induced by X on $\mathcal{B}$. Then

$$F(x) - F(x-0) = \mathbb{P}_X(\{x\}) = \mathbb{P}(X^{-1}(\{x\})).$$

Consequently, if x is a point of continuity of F then $\mathbb{P}_X(\{x\}) = 0$ and if x is a point of discontinuity of F then $\mathbb{P}_X(\{x\})$ equals the jump of F at x.

Proof: Let (x_n) be an increasing sequence in $\mathbb{R}$ converging to x: $\lim x_n = x$. Let $I_n = (x_n, x]$. Then $I_{n+1} \subseteq I_n$



so that I_n is a monotonic

sequence with $\lim_n I_n = \bigcap_n I_n = \{x\}$. Thus by the monotone sequential continuity of $\mathbb{P}_X$, we have

$$\mathbb{P}_X(\{x\}) = \mathbb{P}_X(\lim_n I_n) = \lim_n \mathbb{P}_X(I_n)$$

$$= \lim_n \mathbb{P}_X((x_n, x])$$

$\boxed{(-\infty,x] = (-\infty,x_n] \cup (x_n,x]}$

$$= \lim_n \left(\mathbb{P}_X(-\infty, x] - \mathbb{P}_X(-\infty, x_n]\right)$$

$$= \lim_n F(x) - F(x_n)$$

$$= F(x) - \lim_{n\to\infty} F(x_n)$$

$$= F(x) - F(x-0). \quad \blacksquare$$

<u>Proposition</u>: The set of all discontinuity points of the distribution function of a r.v. is countable.

<u>Proof</u>: Any distribution function is a nondecreasing function $g: \mathbb{R} \to \mathbb{R}$. So it is enough to prove that such a function can have at most countably many discontinuity. Assume on the contrary that g has uncountably many jumps. For any pair of integers $m \in \mathbb{Z}$, $n > 0$, let

$$A_{m,n} = \{x \in \mathbb{R} \mid m \leq x < m+1,\ g(x+0) - g(x-0) \geq 1/n\}.$$

Note that any point of discontinuity of g lies in some $A_{m,n}$. Hence, by the assumption the union

$$\bigcup_{m\in\mathbb{Z},\, n\in\mathbb{Z}^+} A_{m,n}$$ is uncountable. Since $\mathbb{Z}\times\mathbb{Z}^+$ is countable and countable union of countable sets is countable one of the $A_{m,n}$'s must be uncountable say A_{m_0,n_0}. So the interval $[m_0, m_0+1)$ contains infinitely many jumps say $x_1, x_2, \ldots, x_n, \ldots$ so that $g(x_i+0) - g(x_i-0) \geq 1/n_0$.

Thus $g(m_0+1)-g(m_0) \geq \sum_{i=1}^{\infty} g(x_i+0)-g(x_i-0)=+\infty$, a contradiction. This finishes the proof. ∎

Proposition: Let F be the distribution function of r.v. X. If $D=\{x_i \mid i\in I\}$ is the set of all discontinuity points of F and $\mathbb{P}_X(\{x_i\})$ denotes the jump size of F at x_i, then the function

$$F_d(x) = \sum_{i\in I} \mathbb{P}_X(x_i) I_{x_i}(x) = \sum_{x_i\leq x} \mathbb{P}_X(x_i), \text{ where}$$

$$I_{x_i}(x) = \mathbb{1}_{[x_i,\infty)}(x) = \begin{cases} 1 & \text{if } x\in[x_i,\infty) \\ 0 & \text{if } x\notin[x_i,\infty) \end{cases}$$

is non-negative, non decreasing and right continuous with $F_d(-\infty)=0$ and $F_d(\infty)\leq 1$. The set of all discontinuity points of F_d coincides with D and $F_d(x_i)-F_d(x_i-0)=\mathbb{P}_X(\{x_i\})$.

Moreover, $F(x)=F_c(x)+F_d(x)$, where F_c is a non-negative, non-decreasing continuous function with $F_c(-\infty)=0$ and $F_c(+\infty)\leq 1$.

Proof Let $x'<x''$ be two real numbers. Then

$$F_d(x'')-F_d(x')=\mathbb{P}_X(D\cap(x',x''])\leq \mathbb{P}_X((x',x''])=\underbrace{F(x'')-F(x')}_{\geq 0}$$

$\Rightarrow F_d$ is non decreasing.

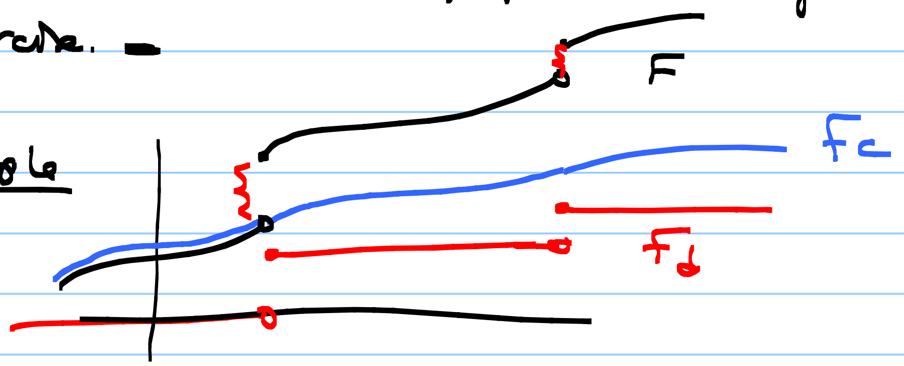
Also $F_d(x)=\mathbb{P}_X(D\cap(-\infty,x])\leq \mathbb{P}_X((-\infty,x])=F(x)$.

In particular,

1) F_d is non decreasing
2) $F_c = F-F_d \geq 0$ is non negative
3) $F_c(x'')-F_c(x') = F(x'')-F_d(x'')-(F(x')-F_d(x'))\geq 0$
$\Rightarrow F_c$ is non decreasing.

The other stated properties are left as an exercise. ∎

Example



$$F = F_d + F_c$$

Proposition: Let F be the distribution function of a real random variable X and $\mathbb{P}_X$ the corresponding induced probability on $\mathcal{B}$. Then for all $a, b \in \mathbb{R}$ such that $a < b$, we have

i) $\mathbb{P}_X((a,b]) = F(b) - F(a)$
ii) $\mathbb{P}_X((a,b)) = F(b-0) - F(a)$
iii) $\mathbb{P}_X([a,b]) = F(b) - F(a-0)$
iv) $\mathbb{P}_X([a,b)) = F(b-0) - F(a-0)$.

Proof: (i) was proved on page 84.

ii) Let (x_n) be an increasing sequence with $\lim x_n = b$, and $x_n \in (a,b)$. Then $(a,b) = \bigcup_{n=1}^{\infty} (a, x_n]$. Hence

$$\mathbb{P}_X((a,b)) = \mathbb{P}_X(\lim_n (a, x_n]) = \lim_n \mathbb{P}_X((a, x_n])$$
$$= \lim_n F(x_n) - F(a)$$
$$= F(b-0) - F(a)$$

The other ones can be proved similarly. ∎

Remark: Note that for any $a \in \mathbb{R}$

1) $\mathbb{P}_X(\{a\}) = \lim_{b \to a^+} \mathbb{P}_X([a,b)) = \lim_{b \to a^+} F(b) - F(a-0)$

$$= F(a) - F(a-0)$$

2) $\mathbb{P}_X((a,\infty)) = \mathbb{P}(\{X > a\}) = \lim_{b \to \infty} \mathbb{P}_X((a,b])$

$$= \lim_{b \to \infty} F(b) - F(a)$$

$$= F(\infty) - F(a)$$

$$= 1 - F(a).$$

3) $\mathbb{P}_X([a,\infty)) = \mathbb{P}(\{X \geq a\}) = \lim_{b \to \infty} \mathbb{P}_X([a,b])$

$$= \lim_{b \to \infty} F(b) - F(a-0)$$

$$= F(\infty) - F(a-0)$$

$$= 1 - F(a-0).$$

Theorem: Let $F(x)$, $x \in \mathbb{R}$, be a function having the following properties

1) $F(x)$ is non-decreasing
2) right continuous
3) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ and $F(\infty) = \lim_{x \to \infty} F(x) = 1$.

Then there exists a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a random variable X such that $F(x) = F_X(x)$.

Proof: From measure theory it is known that if $f(x)$ is a non decreasing, right continuous function on $\mathbb{R}$ then there is a unique measure m on the Borel σ-algebra such that

$$m((a,b]) = f(b) - f(a).$$

We construct $(\Omega, \mathcal{A}, \mathbb{P})$ as follows: let $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}$ and $\mathbb{P} = m$ the measure above.
Then $\mathbb{P}((a,b]) = m((a,b]) = F(b) - F(a)$.

$$\text{Then, } \mathbb{P}(\mathbb{R}) = \mathbb{P}\Big(\lim_{n\to\infty} (-n,n]\Big) = \lim_{n\to\infty} \mathbb{P}((-n,n])$$
$$= \lim_{n\to\infty} F(n) - F(-n) = 1 - 0 = 1.$$

Hence $\mathbb{P}$ is a probability measure.
Define the random variable X on $\Omega = \mathbb{R}$ as $X(\omega) = \omega$, $\omega \in \mathbb{R}$. Then
$\mathbb{P}(X \leq x) = \mathbb{P}(\{\omega / X(\omega) \leq x\}) = \mathbb{P}_X((-\infty, x]) = F_X(x)$.

Moreover,
$$\mathbb{P}(X \leq x) = \mathbb{P}((-\infty, x]) = \lim_{n\to\infty} \mathbb{P}((-n,x]) = \lim_{n\to\infty} F(x) - F(-n)$$
$$= F(x) - F(-\infty)$$
$$= F(x) - 0 = F(x).$$

So $F_X(x) = F(x)$, for all $x \in \mathbb{R}$. ■

<u>Remark</u>: Clearly $\mathbb{P}_X$ and F_X determine each other uniquely.
<u>Question</u>: Does $\mathbb{P}_X$ (or F_X) determine X uniquely?

The answer is no as shown by the example below:
<u>Example</u> Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the probability space of tossing a coin (classical scheme). Define

$$X_1(\omega) = \begin{cases} 1 & \text{if } \omega = h \\ -1 & \text{if } \omega = t \end{cases} \quad \text{and} \quad X_2(\omega) = \begin{cases} -1 & \text{if } \omega = h \\ 1 & \text{if } \omega = t. \end{cases}$$

$$\text{Then } F_{X_1}(x) = F_{X_2}(x) = \begin{cases} 0 & \text{if } x < -1 \\ 1/2 & -1 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

CHAPTER 5: Expectation (Integration) of Random Variables

§5.1 More About Random Variables.

In order to define integrals of random variables first we extend the Borel measure space $(\mathbb{R}, \mathcal{B})$ to $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$, where $\overline{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$ and $\overline{\mathcal{B}} = \sigma(\{-\infty\} \cup \mathcal{B} \cup \{+\infty\})$. It is easy to see that $\overline{\mathcal{B}} = \sigma(\{[-\infty, x] \mid x \in \mathbb{Q}\})$.

Also recall the following properties of the indicator functions:

$\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$ because $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$
$\Rightarrow \mathbb{1}_A(x) + \mathbb{1}_{A^c}(x) = 1$.

$\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B$ if $A \cap B = \emptyset$.

$\mathbb{1}_A \mathbb{1}_B = \mathbb{1}_{A \cap B}$, $\sup(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{1}_{A \cup B}$ and

$\inf(\mathbb{1}_A, \mathbb{1}_B) = \mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$.

Definition: A simple real random variable (srrv) on $(\Omega, \mathcal{A}, \mathbb{P})$ is a real valued function X on Ω of the form

$$X = \sum_{i \in I} x_i \mathbb{1}_{A_i}, \text{ where } \{A_i \mid i \in I\} \text{ is a}$$

finite measurable partition of Ω and $x_i \in \mathbb{R}$, $\forall i \in I$.

Let S be the set of all srrv's on $(\Omega, \mathcal{A}, \mathbb{P})$.

Proposition: Let $X = \sum_{i \in I} x_i \mathbb{1}_{A_i}$ and $Y = \sum_{j \in J} y_j \mathbb{1}_{B_j}$ be in S. Then,

i) S is a vector space with

$$X+Y=\sum_{(i,j)\in I\times J}(x_i+y_j)\,\mathbb{1}_{A_i\cap B_j},$$

$$aX=\sum_{i\in I} a x_i\,\mathbb{1}_{A_i},\quad \forall a\in\mathbb{R}.$$

i) S is an algebra, when $X\cdot Y$ is defined to be

$$XY=\sum_{(i,j)\in I\times J} x_i y_j\,\mathbb{1}_{A_i\cap B_j}.$$

ii) S is a lattice for the natural ordering:
$X\le Y \Rightarrow X(\omega)\le Y(\omega)$, $\forall \omega\in\Omega$, and

$$\sup(X,Y)=\sum_{(i,j)\in I\times J}\sup(x_i,y_j)\,\mathbb{1}_{A_i\cap B_j},\quad\text{and}$$

$$\inf(X,Y)=\sum_{(i,j)\in I\times J}\inf(x_i,y_j)\,\mathbb{1}_{A_i\cap B_j}.$$

Also note that $\sup(-X,-Y)=-\inf(X,Y)$ and
$\sup(X,Y)+\inf(X,Y)=X+Y$

Proof is left as an exercise.

<u>Definition:</u> An extended real random variable (errv) X is an $(\mathcal{A},\overline{\mathcal{B}})$-measurable mapping of Ω into $\overline{\mathbb{R}}$.

<u>Proposition:</u> A function $X:\Omega\to\overline{\mathbb{R}}$ is a errv if and only if $\forall x\in\mathbb{R}\ (x\in\mathbb{Q})\ X^{-1}([-\infty,x])\in\mathcal{A}$.

<u>Proposition:</u> Let $\{X_i\,|\,i\in I\}$ be a countable set of errvs. Then $\inf\{X_i\,|\,i\in I\}$ and $\sup\{X_i\,|\,i\in I\}$ are errvs. If $\{X_n\,|\,n\ge 1\}$ is a sequence of errvs, then $\liminf\{X_n\,|\,n\ge 1\}$ and $\limsup\{X_n\,|\,n\ge 1\}$ are

errvs. Finally if $\{X_n | n \geq 1\}$ converges then $\lim X_n$ is an errv.

$X_i^{-1}([x,\infty])$

Proof: $\{\omega \in \Omega | \inf_i X_i(\omega) \geq x\} = \bigcap_{i \in I} \{\omega | X_i(\omega) \geq x\} \in \mathcal{A}$

and $\{x \in \Omega | \sup_i X_i(\omega) \leq x\} = \bigcap_{i \in I} \{\omega | X_i(\omega) \leq x\} \in \mathcal{A}$

$X_i^{-1}([-\infty, x])$

hold for all $x \in \mathbb{R}$. This implies that $\inf\{X_i\}$ and $\sup\{X_i\}$ are errvs.

If $\{X_n(\omega) | n \geq 1\}$ is a sequence of errvs, the equalities

$\liminf\{X_n\} = \sup_m \{\inf_{n \geq m}\{X_n\}\}$ and

$\limsup\{X_n\} = \inf_m \{\sup_{n \geq m}\{X_n\}\}$ show that the

two limits are errvs. For the final statement, note that if $\{X_n\}$ is convergent then $\lim_n X_n = \limsup_n X_n = \liminf_n X_n$ is an errv.

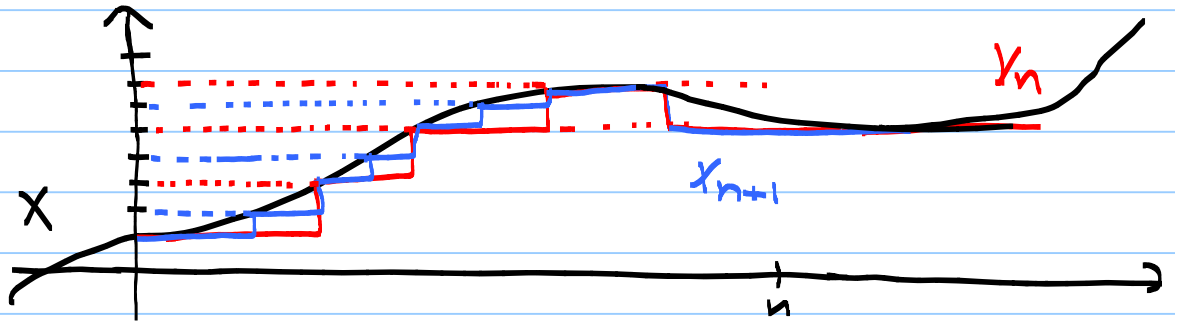
Theorem: $X : \Omega \to \overline{\mathbb{R}}$ is an errv if and only if there is a sequence in S converging pointwise to X. If X is a bounded errv, then one can find a sequence in S converging to X uniformly.

Proof: If $X : \Omega \to \overline{\mathbb{R}}$ is the pointwise limit of a sequence of $\mathcal{A}$-measurable srvs, then by the above proposition so is X.

For the converse statement, let X be a non-negative errv and consider the sequence

Video 33

$$X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{\{\frac{k-1}{2^n} \le X < \frac{k}{2^n}\}} + n \mathbb{1}_{\{X \ge n\}}$$



It is clear that $\{X_n \mid n \ge 1\}$ is an increasing sequence of non-negative srvs converging pointwise to X.
Let X as $X = X^+ - X^-$, where both X^+ and X^- are non negative functions (let $X^+ = \sup\{X, 0\}$ and $X^- = -\inf\{X, 0\}$). Then there are sequences $\{X_n^+\}$ and $\{X_n^-\}$ converging to X^+ and X^- pointwisely. It is clear that $X_n = X_n^+ - X_n^-$ will converge to X pointwisely.
If X is bounded then $\sup_{\omega \in \Omega} |X(\omega) - X_n(\omega)| < \frac{1}{2^n}$ as soon as $n > \sup|X(\omega)|$.
Hence, the convergence is uniform. ∎

Proposition: Let $X_1, \dots, X_n$ and Y be erv's on $(\Omega, \mathcal{A}, \mathbb{P})$ and let $\mathcal{C} := \sigma(X_1, \dots, X_n)$, the smallest sub-σ-algebra generated by $X_1, \dots, X_n$. Then Y is $\mathcal{C}$-measurable if and only if there is a $(\overline{\mathcal{B}}^n, \overline{\mathcal{B}})$-measurable mapping $f: \overline{\mathbb{R}}^n \to \overline{\mathbb{R}}$ such that $Y = f(X_1, X_2, \dots, X_n)$, where $\overline{\mathcal{B}}^n = \overline{\mathcal{B}} \otimes \dots \otimes \overline{\mathcal{B}}$.

Proof: We'll give a proof only for $n=1$. If $Y = f(X)$ for some measurable function $f: \overline{\mathbb{R}} \to \overline{\mathbb{R}}$, then clearly Y is $\sigma(X)$-measurable.
Conversely, assume that Y is $\sigma(X)$-measurable. Let $\{Y_k\}$ be a sequence of srvs converging to

4. Each Y_k has the form $Y_k = \sum_{i=1}^{n} b_i \mathbb{1}_{S_i}$, for some $b_i \in \mathbb{R}$ and $S_i \in \sigma(X)$, $i=1,\dots,n$; $S_i \cap S_j = \phi$ if $i \neq j$ and $\bigcup_{i=1}^{n} S_i = \Omega$.

It is known that $\sigma(X) = \{X^{-1}(B) \mid B \in \mathfrak{B}\}$. Hence, for all $i = 1, \dots, n$, there is some $B_i \in \mathfrak{B}$ s.t. $S_i = X^{-1}(B_i)$. Let $\mathcal{R}(X) = X(\Omega)$. Note that $B_i \cap \mathcal{R}(X)$, $i=1,\dots,n$, are all disjoint: If $x \in B_i \cap \mathcal{R}(X) \cap B_j$, then $\phi \neq X^{-1}(\{x\}) \subseteq X^{-1}(B_i) \cap X^{-1}(B_j) = S_i \cap S_j = \phi$, a contradiction, since $x \in \mathcal{R}(X) = X(\Omega)$.

Define $B_1' \doteq B_1$, $B_2' = B_2 \setminus B_1$, $B_3' = B_3 \setminus (B_1 \cup B_2)$, ..., $B_n' = B_n \setminus (B_1 \cup \dots \cup B_{n-1})$. Note that $B_i' \cap \mathcal{R}(X) = B_i \cap \mathcal{R}(X)$, for all $i = 1, 2, \dots, n$ and B_i''s are disjoint.

Now define

$$f_k \doteq \sum_{i=1}^{n} b_i \mathbb{1}_{B_i'}$$

Then $Y_k(\omega) = \sum_{i=1}^{n} b_i \mathbb{1}_{S_i}(\omega) = \sum_{i=1}^{n} b_i \mathbb{1}_{B_i'}(X(\omega)) = f_k(X(\omega))$.

Define also, $f \doteq \liminf f_k$. Then we have, for all $\omega \in \Omega$,

$Y(\omega) = \lim_{k\to\infty} Y_k(\omega) = \lim_{k\to\infty} f_k(X(\omega)) = f(X(\omega))$, because

$\{f_k(X(\omega)) \mid k \geq 1\}$ is convergent. ∎

§ 5.2. Expectation of Random Variables

Definition. The expectation (or the integral) of a simple random variable $X = \sum_{i=1}^{n} x_i \mathbb{1}_{A_i}$ on $(\Omega, \mathcal{A}, \mathbb{P})$ is the real number

$E(X) = \sum_{i=1}^{n} x_i \mathbb{P}(A_i)$, written as

$E(X) = \int_\Omega X(\omega)\, \mathbb{P}(d\omega) = \int_\Omega X \, d\mathbb{P}.$

Proposition: The expectation E is a mapping of the space S of errus into $\mathbb{R}$ having the following properties:

i) $\forall A \in \mathcal{A}$, $E(\mathbb{1}_A) = \mathbb{P}(A)$, $E(\mathbb{1}) = \mathbb{P}(\Omega) = 1$;
ii) $\forall c \in \mathbb{R}$, $\forall X \in S$, $E(cX) = cE(X)$;
iii) $\forall X, Y \in S$, $E(X+Y) = E(X) + E(Y)$;
iv) $\forall X, Y \in S$, if $X \leq Y$ then $E(X) \leq E(Y)$;
v) If $X_n \in S$, $X \in S$ and (X_n) is an increasing sequence with $\lim X_n = X$ then $\lim E(X_n) = E(X)$.
The same also holds for decreasing sequences.

Proof: (i) and (ii) are easy.
(iii) $X = \sum_i a_i \mathbb{1}_{A_i}$, $Y = \sum_j b_j \mathbb{1}_{B_j}$, $\cup A_i = \Omega = \cup B_j$

$$X + Y = \sum_{i,j} a_i \mathbb{1}_{A_i \cap B_j} + \sum_{i,j} b_j \mathbb{1}_{A_i \cap B_j}$$

$$= \sum_{i,j} (a_i + b_j) \mathbb{1}_{A_i \cap B_j}$$

$$E(X+Y) = \sum_{i,j} a_i + b_j \,\mathbb{P}(A_i \cap B_j)$$

$$= \sum_{i,j} a_i \mathbb{P}(A_i \cap B_j) + \sum_{i,j} b_j \mathbb{P}(A_i \cap B_j)$$

$$= \sum_i a_i \mathbb{P}(A_i \cap \Omega) + \sum_j b_j \mathbb{P}(\Omega \cap B_j)$$

$$= E(X) + E(Y).$$

(iv) If $X \leq Y$ and $X = \sum x_i \mathbb{1}_{A_i}$, $Y = \sum y_j \mathbb{1}_{B_j}$, then

$$Y - X = \sum_{i,j} (y_j - x_i) \mathbb{1}_{A_i \cap B_j}, \quad y_j - x_i \geq 0, \text{ if } A_i \cap B_j \neq \emptyset.$$

Also note that if $X \geq 0$ then $E(X) \geq 0$.

Video 34

So, writing $Y = X + (Y-X)$ we get
$$E(Y) = E(X) + E(Y-X) \geq E(X), \text{ since } Y - X \geq 0.$$

v) Let (X_n) be a decreasing sequence with $\lim X_n = X$. Then $(X_n - X)$ is a non negative decreasing sequence with $\lim (X_n - X) = 0$. Let k be the maximum value of $X_n - X$. So, $0 \leq \tilde{X}_n \leq k$ and hence, for any fixed $\epsilon > 0$, $0 \leq \tilde{X}_n \leq k \, \mathbb{1}_{(\tilde{X}_n > \epsilon)} + \epsilon$, $\tilde{X}_n = X_n - X$

Note that the sequence of measurable sets $(\tilde{X}_n > \epsilon)$ is decreasing with $\lim (\tilde{X}_n > \epsilon) = \bigcap_n (\tilde{X}_n > \epsilon) = \phi$.
So, $0 \leq E(\tilde{X}_n) \leq k \mathbb{P}(\tilde{X}_n > \epsilon) + \epsilon$ and taking limits as $n \to \infty$ we get
$$0 \leq \lim E(\tilde{X}_n) \leq k \lim \mathbb{P}(\tilde{X}_n > \epsilon) + \epsilon$$
$$= k \, \mathbb{P}(\underbrace{\lim (\tilde{X}_n > \epsilon)}_{= \phi}) + \epsilon$$
$$= \epsilon$$
Since, ϵ were arbitrary, we see that
$$\lim E(\tilde{X}_n) = 0.$$
$$0 = \lim E(\tilde{X}_n) = \lim_n E(X_n - X) = \lim (E(X_n) - E(X))$$
$$= \lim E(X_n) - E(X).$$
$$\Rightarrow E(X) = \lim E(X_n). \quad \blacksquare$$

Definition: Let X be a non-negative error on $(\Omega, \mathcal{A}, \mathbb{P})$. The expectation or the integral of X is defined as $E(X) = \lim_n E(X_n)$, for any increasing sequence of nonnegative srvs (X_n) converging to X.

Remark: $E(X)$ is well defined. To see this let (X_n) and (Y_n) be two increasing sequences of nonnegative srvs converging to X.
must show: $\lim E(X_n) = \lim E(Y_n)$.

Proposition. Let S^+ be the class of all nonnegative srvs on $(\Omega, \mathcal{A}, \mathbb{P})$. If $\{X_m \mid m\in\mathbb{N}\}$ and $\{Y_n \mid n\in\mathbb{N}\}$ are two increasing sequences in S^+ and if $\lim_m X_m \le \lim_n Y_n$, then $\lim_m E(X_m) \le \lim_n E(Y_n)$.
Consequently, for a non negative erv the definition of $E(X)$ is independent of the chosen particular sequence (X_m) converging to X.

Proof $\inf(X_m, Y_n) = X_m \mathbb{1}_{(X_m \le Y_n)} + Y_n \mathbb{1}_{(X_m > Y_n)}$

and hence $\lim_{n\to\infty} \inf(X_m, Y_n) = X_m \mathbb{1}_{\Omega} + Y \mathbb{1}_{\emptyset}$
$= X_m$, where $Y = \lim Y_n$.

Also, $Y_n \ge \inf(X_m, Y_n)$ and thus $E(Y_n) \ge E(\inf(X_m, Y_n))$ and $\lim_{n\to\infty} E(Y_n) \ge \lim_{n\to\infty} E(\inf(X_m, Y_n)) = E(X_m)$, $\forall m$.

So, $\lim_{n\to\infty} E(Y_n) \ge \lim_m E(X_m)$.

For the final statement if (X_n) and (Y_n) are two sequences in S^+ converging to X then
Then $\lim X_n = \lim Y_n$ and hence $\lim X_n \le \lim Y_n$
$\Rightarrow \lim E(X_n) \le \lim E(Y_n)$. Similarly, $\lim Y_n \le \lim X_n$
$\Rightarrow \lim E(Y_n) \le \lim E(X_n)$. So
$\lim E(X_n) = \lim E(Y_n)$. ∎

Proposition: Let L^+ be the class of all nonnegative ervs on $(\Omega, \mathcal{A}, \mathbb{P})$. Then
i) $X \in L^+$ then $0 \le E(X) \le \infty$.
ii) $X \in L^+$, $c \ge 0$ then $cX \in L^+$ and $E(cX) = cE(X)$.
iii) $X, Y \in L^+$ then $X+Y \in L^+$ and $E(X+Y) = E(X) + E(Y)$.
iv) $X, Y \in L^+$ with $X \le Y$ then $E(X) \le E(Y)$.
v) If (X_n) is an increasing sequence with $\lim X_n$ in L^+ then $\lim X_n \in L^+$ and $E(\lim X_n) = \lim E(X_n)$.

Remark (v) is valid for decreasing sequences also and it can be deduced from Monotone Convergence Theorem or the Dominated Convergence Theorem, which will be proved later.

Proof: (i), (ii) are easy. (iv) is proved in the previous proposition.

For (iii) note that if (X_n) and (Y_n) are increasing sequences in S^+ converging to X and Y, respectively, then (X_n+Y_n) is also an increasing sequence converging to $X+Y$ in S^+. Then

$$E(X+Y) = \lim_n E(X_n+Y_n) = \lim E(X_n)+E(Y_n)$$
$$= \lim E(X_n) + \lim E(Y_n).$$

To prove (v) let (X_n) be an increasing sequence in L^+ and let $\{Y_{n,m} \mid m \in \mathbb{N}\}$ be an increasing sequence in S^+ converging to X_n.

Put $Z_m = \sup_{n \leq m} Y_{m,n}$. Since $Y_{m,n} \leq Y_{m+1,n}$ and

$$Z_m = \sup_n \{Y_{m,n} \mid n \leq m\} \leq \sup_n \{Y_{m+1,n} \mid n \leq m\}$$
$$\leq \sup_n \{Y_{m+1,n} \mid n \leq m+1\} = Z_{m+1},$$

(Z_m) is an increasing sequence in S^+. Moreover, as $X_1 \leq X_2 \leq \cdots \leq X_n \leq \cdots$ and $Y_{m,1} \leq X_1$, $Y_{m,2} \leq X_2, \ldots$, $Y_{m,m} \leq X_m$ we have

$$Z_m = \sup_n \{Y_{m,n} \mid n \leq m\} \leq \sup(X_1, \ldots, X_m) = X_m.$$

Thus, $X_n = \lim_{m\to\infty} Y_{m,n} \leq \lim_{m\to\infty} Z_m \leq \underline{\lim} X_m$. Hence

$$\lim X_n \leq \lim_{m\to\infty} Z_m \leq \lim X_m \in L^+.$$

Similarly, $E(Y_{m,n}) \le E(Z_m) \le E(X_n)$ and by the previous proposition

$$E(X_n) = \lim_m E(Y_{m,n}) \le \lim_m E(Z_m)$$
$$= E(\lim Z_m)$$
$$= E(\lim X_n)$$
$$= \lim E(X_n)$$

Therefore $\lim E(X_n) = E(\lim X_n)$. ∎

<u>Definition:</u> A non negative ernv X is said to be integrable if $E(X) < \infty$. An arbitrary X is said to quasi-integrable if either $E(X^+)$ or $E(X^-)$ is finite. In this case, $E(X)$ is defined as $E(X) = E(X^+) - E(X^-)$. We say that an ernv X is integrable if $E(X)$ is finite, i.e, if X^+ and X^- are integrable, or equivalently, $|X|$ is integrable. In this case, we say that $E(X)$, the expectation of X exists. Whenever, $E(X)$ is defined we write

$$E(X) = \int_\Omega X(\omega)\, \mathbb{P}(d\omega) = \int_\Omega X \, d\mathbb{P}$$ and read

this as X is $\mathbb{P}$-integrable.

<u>Proposition:</u> Let X and Y be ernvs. Then

i) If X is integrable then, $\mathbb{P}(\{X = \pm\infty\}) = 0$;
ii) $E(cX) = cE(X)$, $\forall c \in \mathbb{R}$,
iii) $E(X+Y) = E(X) + E(Y)$, if $X+Y$ is defined and if $X^+ + Y^+$ or $X^- + Y^-$ is integrable.
iv) $X \le Y \Rightarrow E(X) \le E(Y)$.

<u>Proof:</u> i) X is integrable if and only if $E(|X|) < \infty$. Consider the sequence (X_n) converging to $|X|$ defined earlier as

$$X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{\{\frac{k-1}{2^n} \leq |X| < \frac{k}{2^n}\}} + n \mathbb{1}_{\{|X| \geq n\}}.$$

So $\mathbb{E}(X_n) = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{P}(\{\frac{k-1}{2^n} \leq |X| < \frac{k}{2^n}\}) + n \mathbb{P}(\{|X| \geq n\})$.

Since $E(X_n)$ is an increasing sequence converging to $E(X) < \infty$, then $n\mathbb{P}(\{|X| \geq n\})$ is a bounded sequence. This is possible only if $\lim_n \mathbb{P}(\{|X| \geq n\}) = 0$. Finally, $\{|X| = \infty\} = \bigcap_{n=1}^{\infty} \{|X| \geq n\}$ and $(\{|X| \geq n\})$ is a monotone (decreasing) sequence

$$\mathbb{P}(\{|X| = \infty\}) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \{|X| \geq n\}\right) = \mathbb{P}(\lim \{|X| \geq n\})$$
$$= \lim \mathbb{P}(\{|X| \geq n\})$$
$$= 0.$$

ii) Obvious.

iii) The decomposition $(X^+ - X^-) + (Y^+ - Y^-) = (X+Y)^+ - (X+Y)^-$ gives $X^+ + Y^+ + (X+Y)^- = (X+Y)^+ + X^- + Y^-$, where all terms are non negative. Then by the previous proposition

$$E(X^+) + E(Y^+) + E((X+Y)^-) = E((X+Y)^+) + E(X^-) + E(Y^-)$$

and this implies

$$E(X+Y) = E((X+Y)^+) - E((X+Y)^-)$$
$$= E(X^+) - E(X^-) + E(Y^+) - E(Y^-)$$
$$= E(X) + E(Y).$$

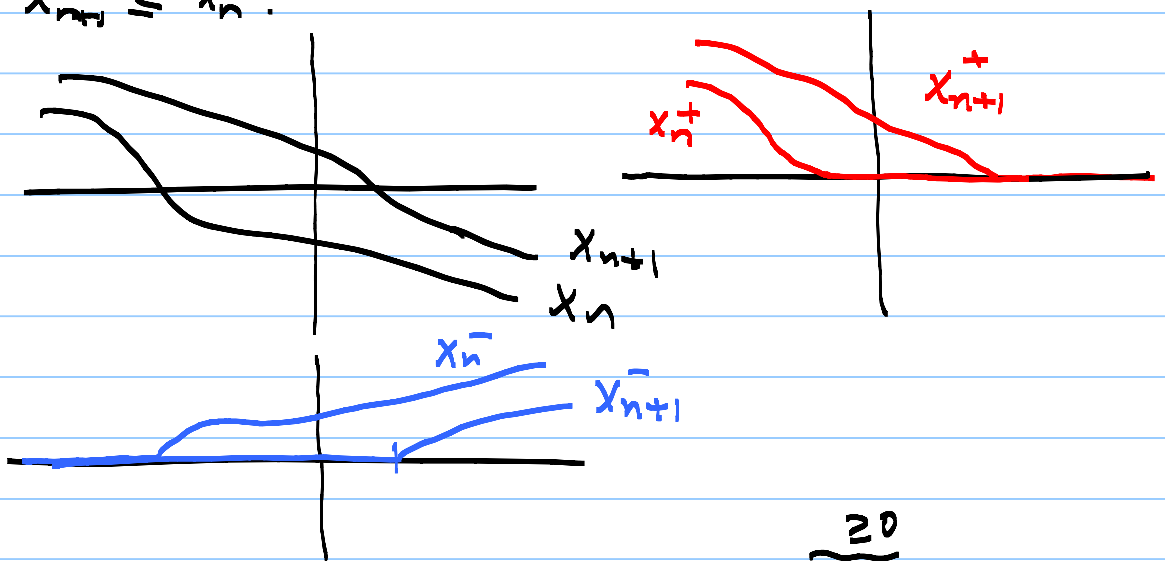
iv) If $X \leq Y$ i.e. $X^+ - X^- \leq Y^+ - Y^-$ then $X^+ + Y^- \leq Y^+ + X^-$. So again by the previous proposition we see that $E(X^+) + E(Y^-) \leq E(Y^+) + E(X^-)$ and therefore,

$$E(X) = E(X^+) - E(X^-) \leq E(Y^+) - E(Y^-) = E(Y). \quad \blacksquare$$

Theorem (Monotone Convergence Theorem)

Let $\{X_n \mid n \in \mathbb{N}\}$ be a monotone sequence of ernvs. If $X_n \uparrow X$ and $E(X_0^-) < \infty$, then $E(X_n) \uparrow E(X)$ and if $X_n \downarrow X$ and $E(X_0^+) < \infty$, then $E(X_n) \downarrow E(X)$.

Proof: Let $X_n \uparrow X$ and $E(X_0^-) < \infty$. Then for any n, $E(X_n)$ is well defined and $X_n^+ \leq X_{n+1}^+$ and $X_{n+1}^- \leq X_n^-$.



In particular, $X_n^+ - X_n^- + X_0^- = X_n^+ + \underbrace{(X_0^- - X_n^-)}_{\geq 0} \geq 0$ and thus, $0 \leq X_n + X_0^- \uparrow X + X_0^-$. So

$$E(X_n + X_0^-) = E(X_n) + E(X_0^-) \uparrow E(X + X_0^-) = E(X) + E(X_0^-).$$

Hence, $\lim E(X_n) = E(X)$.
The other part is similar. ■

Remark: In the above proof the assumption that $E(X_0^-) < \infty$ can be relaxed as $E(X_m^-) < \infty$ for some m.

Corollary If (X_n) is an increasing (resp. decreasing) sequence of ernvs and if Y is an ernv such that $E(Y^-) < \infty$ (resp. $E(Y^+) < \infty$) and

Video 36

$X_n \geq Y$ (resp. $X_n \leq Y$), for all n, then $\lim_{n\to\infty} X_n$ is quasi-integrable and $E(X_n) \uparrow E(\lim_n X_n)$ (resp. $\downarrow$).

Proof: left as an exercise!

Theorem (The Lebesgue-Fatou Lemma and the Dominated Convergence Theorem)
Let $\{X_n \mid n\in\mathbb{N}\}$ be a sequence of errvs and let Y and Z be integrable errvs. Then
$X_n \leq Y$ $\forall n$, then $\limsup E(X_n) \leq E(\limsup X_n)$
$X_n \geq Z$ $\forall n$, then $E(\liminf X_n) \leq \liminf E(X_n)$.
(This is Lebesgue-Fatou Lemma)

In particular, if $\{X_n \mid n\in\mathbb{N}\}$ converges and if there is an integrable positive errv U such that $|X_n| \leq U$ for all n, then

$$E(\lim_n X_n) = \lim_n E(X_n)$$

(This is the Dominated Conv. Thm).

Proof: Suppose $X_n \leq Y$. Then $\sup_{n\geq m} X_n \leq Y$, for any m and hence
$(\sup_{n\geq m} X_n)^+ \leq Y^+$, where Y^+ is also integrable.

$(\sup_{n\geq m} X_n)^+$ is an decreasing sequence in m, and thus $(\sup_{n\geq m} X_n)$ is quasi integrable. So by again

the previous corollary $E(\sup_{n\geq m}) \downarrow E(\limsup X_n)$.

On the other hand, since $\sup_{n\geq m} X_n \geq X_m$ we have

$E(\sup_{n\geq m} X_n) \geq E(X_m)$, for all m. Hence,

$$E(\sup_{n\geq m} X_n) \geq E(\sup_{n\geq k\geq m} X_n) \geq E(X_k), \text{ for any } k\geq m.$$

$$\Rightarrow E(\sup_{n\geq m} X_n) \geq \sup_{k\geq m} E(X_k) = \underline{\sup_{n\geq m} E(X_n)}, \text{ for all } m.$$

This implies that

$$\limsup E(X_n) = \lim_m \sup_{n\geq m} E(X_n) \leq \lim_m E(\sup_{n\geq m} X_n) = E(\limsup X_n)$$

This proves the first statement. The proof of the statement $X_n \geq Z \Rightarrow E(\liminf X_n) \leq \liminf E(X_n)$ is similar.

For the last statement, if $|X_n| \leq U$ and U is integrable, we have $-U \leq X_n \leq U$, $\forall n$ and then

$$E(\liminf X_n) \leq \liminf E(X_n) \leq \limsup E(X_n) \leq E(\limsup X_n).$$

However, since $\{X_n\}$ is convergent, then $\liminf X_n = \lim X_n = \limsup X_n$ and all the inequalities above are indeed equalities. This implies that $\lim E(X_n)$ exists and

$$\lim E(X_n) = E(\limsup X_n) = E(\lim X_n). \blacksquare$$

<u>Definition</u>: Let X be a quasi-integrable erv on $(\Omega, \mathcal{A}, \mathbb{P})$ and let $A\in\mathcal{A}$. Then the integral defined as

$$\int_A X\, d\mathbb{P} \doteq \int_\Omega \mathbb{1}_A X\, d\mathbb{P} \quad \text{is called}$$

the (indefinite) integral of X over A.
??

<u>Proposition</u>: For a quasi-integrable erv X on $(\Omega, \mathcal{A}, \mathbb{P})$ the following properties hold:

i) $\forall A \in \mathcal{A}$, $E(X) = \int_A X\, d\mathbb{P} + \int_{A^c} X\, d\mathbb{P}$

ii) If $\{A_i \mid i\in I\}$ is a countable disjoint class in $\mathcal{A}$, then
$$\int_{\bigcup_{i\in I} A_i} X\,d\mathbb{P} = \sum_{i\in I}\int_{A_i} X\,d\mathbb{P}$$

iii) If $\{A_n \mid n\geq 1\}$ is an increasing (resp. decreasing) sequence in $\mathcal{A}$ s.t. $\bigcup_{n=1}^{\infty} A_n = A$ (resp. $\bigcap_{n=1}^{\infty} A_n = A$), then
$$\lim_{n\to\infty}\int_{A_n} X\,d\mathbb{P} = \int_A X\,d\mathbb{P}.$$

iv) If $X\geq 0$, then for all $A_1, A_2\in\mathcal{A}$ with $A_1\subseteq A_2$ then
$$\int_{A_1} X\,d\mathbb{P} \leq \int_{A_2} X\,d\mathbb{P}.$$

v) If $A\in\mathcal{A}$ is negligible, then $\int_A X\,d\mathbb{P}=0$.

__Proof:__ (iv) follows from the previous results since $X\mathbb{1}_{A_1}\leq X\mathbb{1}_{A_2}$ on Ω and thus
$$\int_{A_1} X\,d\mathbb{P} = \int_\Omega X\mathbb{1}_{A_1}\,d\mathbb{P} \leq \int_\Omega X\mathbb{1}_{A_2}\,d\mathbb{P} = \int_{A_2} X\,d\mathbb{P}.$$

For the other parts first assume that $\mathbb{E}(X^-)<\infty$ (the other case $\mathbb{E}(X^+)<\infty$ is similar).

Note that for any $A\in\mathcal{A}$
$$\int_A X\,d\mathbb{P} = \int_A X^+\,d\mathbb{P} - \int_A X^-\,d\mathbb{P} \quad\text{and}\quad \int_A X^-\,d\mathbb{P}\leq\int_\Omega X^-\,d\mathbb{P}.$$

i) $\mathbb{E}(X) = \int_\Omega (\mathbb{1}_A + \mathbb{1}_{A^c})\,X\,d\mathbb{P} = \int_\Omega (\mathbb{1}_A X + \mathbb{1}_{A^c}X)\,d\mathbb{P}$

$$= \int_\Omega \mathbb{1}_A X \, d\mathbb{P} + \int_\Omega \mathbb{1}_{A^c} X \, d\mathbb{P}$$

$$= \int_A X \, d\mathbb{P} + \int_{A^c} X \, d\mathbb{P}, \quad \llcorner\!\lrcorner$$

Part (iii) of the proposition before Monotone Conv. Thm.

i) If I is finite, we have

$$\int_{\bigcup_{i \in I} A_i} X \, d\mathbb{P} = \int_\Omega \mathbb{1}_{\bigcup_i A_i} X \, d\mathbb{P} = \int_\Omega \Big(\sum_i \mathbb{1}_{A_i} X\Big) d\mathbb{P}$$

$$= \sum_i \int_\Omega \mathbb{1}_{A_i} X \, d\mathbb{P} = \sum_i \int_{A_i} X \, d\mathbb{P}.$$

Now let $\{A_i \mid i \in I\}$ be a countable (non-finite) disjoint class in $\mathcal{A}$. Then

$$\int_{\bigcup_{i=1}^\infty A_i} X \, d\mathbb{P} = \int_\Omega \mathbb{1}_{\bigcup_{i=1}^\infty A_i} X \, d\mathbb{P} = \int_\Omega \sum_{i=1}^\infty \mathbb{1}_{A_i} X^+ \, d\mathbb{P} - \int_\Omega \sum_{i=1}^\infty \mathbb{1}_{A_i} X^- \, d\mathbb{P}$$

$$= \int_\Omega \lim_n \sum_{i=1}^n \mathbb{1}_{A_i} X^+ \, d\mathbb{P} - \int_\Omega \lim_n \sum_{i=1}^\infty \mathbb{1}_{A_i} X^- \, d\mathbb{P}$$

Since both $\{\sum_{i=1}^n \mathbb{1}_{A_i} X^+ \mid n \geq 1\}$ and $\{\sum_{i=1}^n \mathbb{1}_{A_i} X^- \mid n \geq 1\}$

are both increasing sequences in $\mathcal{L}^+$, by a proposition proved earlier we may take the limit outside the integrals and have

$$\int_{\bigcup_{i=1}^\infty A_i} X \, d\mathbb{P} = \lim_n \sum_{i=1}^n \underbrace{\int_\Omega \mathbb{1}_{A_i} X^+ \, dA} - \lim_n \sum_{i=1}^n \underbrace{\int_\Omega \mathbb{1}_{A_i} X^- \, d\mathbb{P}}$$

$$= \lim_n \sum_{i=1}^n \int_\Omega \mathbb{1}_{A_i} (X^+ - X^-) \, dA$$

$$= \lim_n \sum_{i=1}^n \int_{A_i} X \, d\mathbb{P} = \sum_{i=1}^\infty \int_{A_i} X \, d\mathbb{P}.$$

Video 37

v) We have $\int_A X \, d\mathbb{P} = \int_A X^+ \, d\mathbb{P} - \int_A X^- \, d\mathbb{P}$. Choose (X_n^+) and (X_n^-) in S^+ s.t. $X_n^+ \uparrow X^+$ and $X_n^- \uparrow X_n^-$.

Then $\mathbb{1}_A X_n^+ \uparrow \mathbb{1}_A X^+$ and $\mathbb{1}_A X_n^- \uparrow \mathbb{1} X^-$.

We may assume that $X_n^{\overset{(-)}{+}} = \sum_{k=1}^{N} \mathbb{1}_{A_k} x_k$, where $\{A_1, \dots, A_N\}$ is a partition of Ω so that

$$\mathbb{1}_A X^{\overset{(-)}{+}} = \sum_{k=1}^{N} \mathbb{1}_{A \cap A_k} x_k \text{, so that}$$

$$\int_A X_n^{\overset{(-)}{+}} d\mathbb{P} = \int_\Omega \mathbb{1}_A X_n^{\overset{(-)}{+}} d\mathbb{P} = \sum_{k=1}^{N} \mathbb{P}(A \cap A_k) x_k = 0$$

since $0 \leq \mathbb{P}(A \cap A_k) \leq \mathbb{P}(A) = 0$.

Hence $\int_A X \, d\mathbb{P} = \int_A X^+ d\mathbb{P} - \int_A X^- d\mathbb{P} = 0$. ■

<u>Proposition</u> (Term by Term Integration)
Let $\{X_n \mid n \in \mathbb{Z}^+\}$ be a sequence of integrable errors on $(\Omega, \mathcal{A}, \mathbb{P})$ and let

$$\sum_{n=1}^{\infty} E(|X_n|) < +\infty.$$

Then $\sum_{n=1}^{\infty} |X_n| < +\infty$ with probability 1 (i.e. almost surely) so that $\sum_{n=1}^{\infty} X_n$ converges with probability 1 and $\int_\Omega \left(\sum_{n=1}^{\infty} X_n\right) d\mathbb{P} = \sum_{n=1}^{\infty} \int_\Omega X_n \, d\mathbb{P}$.

Proof: Since $\sum_{n=1}^{m} |X_n| \uparrow \sum_{n=1}^{\infty} |X_n|$ we see that (Prop. 5.2.5) $\int_{\Omega} \sum_{n=1}^{m} |X_n| \, d\mathbb{P} = \sum_{n=1}^{m} \int_{\Omega} |X_n| \, d\mathbb{P} \uparrow$

$$\sum_{n=1}^{\infty} \int_{\Omega} |X_n| \, d\mathbb{P} = \int_{\Omega} \overbrace{\left(\sum_{n=1}^{\infty} |X_n|\right) d\mathbb{P} < \infty}.$$

Therefore, by Prop 5.2.7, $\sum_{n=1}^{\infty} |X_n| < \infty$ and consequently $\sum_{n=1}^{\infty} X_n$ converges with probability 1.

On the other hand, $\left|\sum_{n=1}^{m} X_n\right| \leq \sum_{n=1}^{m} |X_n| \leq \sum_{n=1}^{\infty} |X_n| < \infty$ so that Thm 5.2.10. implies that

$$\int_{\Omega} \left(\sum_{n=1}^{\infty} X_n\right) d\mathbb{P} = \sum_{n=1}^{\infty} \int_{\Omega} X_n \, d\mathbb{P}.$$

Proposition: Let X be a non-negative rv, then $X = 0$ a.s. iff $E(X) = 0$.

Proof: Let $A = \{X = 0\}$, then $A^c = \{X > 0\}$, where $\Omega = A \cup A^c$.

Now, $E(X) = \int_{\Omega} X \, d\mathbb{P} = \int_{A} X \, d\mathbb{P} + \int_{A^c} X \, d\mathbb{P} = \int_{A^c} X \, d\mathbb{P}$

Let $A_n = \{X \geq 1/n\}$. Then $A_n \subseteq A^c$ and then

$$E(X) = \int_{A^c} X \, d\mathbb{P} \geq \int_{A_n} X \, d\mathbb{P} \geq \int_{A_n} 1/n \, d\mathbb{P} = \frac{1}{n} \mathbb{P}(A_n) \geq 0.$$

Now, if $E(X) = 0$, then $1/n \, \mathbb{P}(A_n) = 0 \;\; \forall n$, so that $\mathbb{P}(A_n) = 0$. Also, $A_n \uparrow A^c$ so that $0 = \mathbb{P}(A_n) \uparrow \mathbb{P}(A^c) \Rightarrow \mathbb{P}(A^c) = 0. \Rightarrow X = 0$ a.s.

Conversely, if $X=0$ a.s. then $\mathbb{P}(A^c)=0$ and hence, $E(X)=\int_{A^c} X\, d\mathbb{P}=0$. ∎

Ex $f:\mathbb{R}\to\mathbb{R}$, $f(x)=\begin{cases}0 & \text{if } x\notin\mathbb{Q}\\ 1 & \text{if } x\in\mathbb{Q}.\end{cases}$

$A=\{x\in\mathbb{R} \mid f(x)=0\}=\mathbb{R}\setminus\mathbb{Q}$
$A^c=\{x\in\mathbb{R} \mid f(x)>0\}=\mathbb{Q}$

$\mathbb{P}(A^c)=\mathbb{P}(\mathbb{Q})=0$ (in Lebesgue measure)
$\Rightarrow f=0$ a.s. and hence, $\int_{\mathbb{R}} f\, d\mathbb{P}=0$.

$(\Omega,\mathcal{A},\mathbb{P})=(\mathbb{R},\mathcal{L},m)$.

Corollary If $X=Y$ a.s. and X or Y is integrable then so is the other one and $E(X)=E(Y)$.

Proof: Just apply the above proposition to $|X-Y|$.

Let $X=Y$ a.s.. Then $|X-Y|=0$ a.s. and thus $|E((X-Y))|\leq E(|X-Y|)=0 \Rightarrow E(X-Y)=0$.

Now, $|Y|\leq|X|+|Y-X|$ and $|X|\leq|Y|+|Y-X|$ so that if $|X|$ or $|Y|$ is integrable then so is the other one. So both $|X|$ and $|Y|$ are integrable. Finally, since $E(X-Y)=0$ we have $0=E(X-Y)=E(X)-E(Y)\Rightarrow E(X)=E(Y)$. ∎

Corollary Let X be an integrable r.v. Then $X = 0$ a.s. iff $\forall A \in \mathcal{A}$, $\int_A X \, d\mathbb{P} = 0$.

Proof: Let $X = 0$ a.s. Then $|X| = 0$ a.s. and for any $A \in \mathcal{A}$

$$\left| \int_A X \, d\mathbb{P} \right| = \left| \int_\Omega \mathbb{1}_A X \, d\mathbb{P} \right| \leq \int_\Omega \mathbb{1}_A |X| \, d\mathbb{P} = \int_A |X| \, d\mathbb{P} = 0$$

by the above proposition. Now conversely, assume that $\int_A X \, d\mathbb{P} = 0 \quad \forall A \in \mathcal{A}$.

Let $A^+ = \{X > 0\}$ and $A^- = \{X < 0\}$. Then

$\mathbb{E}(\mathbb{1}_{A^+} X) = \int \mathbb{1}_{A^+} X \, d\mathbb{P} = \int_{A^+} X \, d\mathbb{P} = 0$. Similarly,

$\mathbb{E}(\mathbb{1}_{A^-} X) = 0$. So by the above proposition

$\mathbb{1}_{A^+} X = 0 = \mathbb{1}_{A^-} X$ a.s. Hence, $X = 0$ a.s. ∎

Proposition: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(\Omega', \mathcal{A}')$ a measurable space, X an $(\mathcal{A}, \mathcal{A}')$ measurable mapping of Ω into Ω' and $\mathbb{P}'$ the probability measure induced by X on $\mathcal{A}'$. If Y is any $\mathbb{P}'$-integrable r.v. on $(\Omega', \mathcal{A}', \mathbb{P}')$ then $Z = Y \circ X$ is a $\mathbb{P}$-integrable r.v. on $(\Omega, \mathcal{A}, \mathbb{P})$ and

$$\int_\Omega Z \, d\mathbb{P} = \int_{\Omega'} Y \, d\mathbb{P}'.$$

Proof: If the above equality holds for $Z^+ = Y^+ \circ X$

Video 3 P

and $Z^- = Y^- \circ X$, then by additivity it holds for all Y. Hence we may assume that $Y \geq 0$. Let $\{Y_n\}$ be a sequence of srrv with $\{Y_n\} \uparrow Y$ and $Y_n \geq 0$ $\forall n$. Then $\{Z_n = Y_n \circ X\}$ is another sequence of srrv with $\{Z_n\} \uparrow Z$.

Now if

$$\int_\Omega Z_n \, d\mathbb{P} = \int_{\Omega'} Y_n \, d\mathbb{P}', \text{ for all } n, \text{ then taking}$$

limit we obtain the result. So it is enough to prove for $Y = Y_n$. So let $Y = \sum_{i=1}^k y_i \mathbb{1}_{A_i'}$ be a s.r.r.v. on $(\Omega', \mathcal{A}', \mathbb{P}')$.

Then $Z = Y \circ X = \sum_{i=1}^k y_i \mathbb{1}_{A_i}$, where $A_i = X^{-1}(A_i')$ $\forall i$.

Finally,
$$\int_\Omega Z \, d\mathbb{P} = \sum_{i=1}^k y_i \mathbb{P}(A_i) = \sum_{i=1}^k y_i \, \mathbb{P}(X^{-1}(A_i'))$$
$$= \sum_{i=1}^k y_i \, \mathbb{P}'(A_i')$$
$$= \int_{\Omega'} Y \, d\mathbb{P}'$$

■

<u>Definition:</u> Let X be an $\mathbb{R}^n$-valued random variable on $(\Omega, \mathcal{A}, \mathbb{P})$, $\mathbb{P}'$ the probability measure induced by X on $\mathcal{B}^n$, F the corresponding distribution function and g be a $(\overline{\mathcal{B}}^n, \overline{\mathcal{B}})$-measurable $\mathbb{P}'$-integrable mapping of $\overline{\mathbb{R}}^n$ into $\overline{\mathbb{R}}$. Then the integral of g with respect to F is defined by

$$\int_{\mathbb{R}^n} g(x) \, dF(x) \doteq \int_{\mathbb{R}^n} g(x) \, \mathbb{P}'(dx) = \int_\Omega g(X(\omega)) \, \mathbb{P}(d\omega).$$

This integral is called the expectation or the mean of g. It is also called the Lebesgue Stieltjes integral of g with respect to F.

Remark. i) Suppose X is a discrete rv with values $\{x_i \mid i \in I\} \subseteq \mathbb{R}^n$, where I is countable, and put $p_i = \mathbb{P}(\{X = x_i\})$. Then the above definition reduces to

$$E(g) = \sum_{i \in I} g(x_i)\, p_i \qquad g : I \longrightarrow \mathbb{R}^n$$

$$(\Omega, \mathcal{A}, \mathbb{P}) = (I, \mathcal{A}, \mathbb{P}) \qquad \left(\sum_{i \in I} \mathbb{P}(i) = 1\right)$$

ii) If F is absolutely continuous with a density function f, then

$$E(g(x)) = \int_{\mathbb{R}^n} g(x) f(x)\, dx,$$ where the integral is a Lebesgue integral.

Definition: A rrv is said to be centered if its mean (or expectation) is zero.

§5.3 Lp–Spaces: Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A rrv X on $(\Omega, \mathcal{A}, \mathbb{P})$ is a measurable function of Ω into $\mathbb{R}$ with domain $D(X) \in \mathcal{A}$ so that $\mathbb{P}(D(X)) = 1$. Two rrvs X and Y are said to be a.s. (or a.e.) equal if

$$\mathbb{P}(\{\omega \in D(X) \cap D(Y) \mid X(\omega) \neq Y(\omega)\}) = 0.$$

In this case, we also say that X and Y are equivalent.

If X is a r.v. define it as zero on $D(X)^c$ and the integral can be now computed on Ω.

Definition:

- $L_0(\Omega, \mathcal{A}, \mathbb{P})$ the space of equivalence classes of all r.v.s on $(\Omega, \mathcal{A}, \mathbb{P})$.

- $L_p(\Omega, \mathcal{A}, \mathbb{P})$, $p \in [1, \infty)$, the space of equivalence classes of all r.v.s on $(\Omega, \mathcal{A}, \mathbb{P})$ such that

$$\int_\Omega |X|^p \, d\mathbb{P} < +\infty.$$

- $L_\infty(\Omega, \mathcal{A}, \mathbb{P})$ the space of equivalence classes of all r.v.s on $(\Omega, \mathcal{A}, \mathbb{P})$ such that

$$\operatorname{ess\,sup} |X| \doteq \sup\{x \in \mathbb{R}^+ \mid \mathbb{P}(|X| \geq x) > 0\} < \infty.$$

Ex $f: [0,1] \to \mathbb{R}$, $f(x) = \begin{cases} x & x \neq 1/n \quad n = 1, 2, \ldots \\ n & \text{if } x = 1/n \end{cases}$

f is not bounded, but $f \in L_\infty$.

The above spaces are real vector spaces and have lattice structure with the following operations:

$$[X] + [Y] = [X + Y], \quad c[X] = [cX].$$

$$\sup([X], [Y]) = [\sup(X, Y)] \text{ and}$$

$$\inf([X], [Y]) = [\inf(X, Y)].$$

Theorem (Jensen Inequality)

Let X and Y be integrable r.v.s and let φ be a convex (resp. concave) continuous function of $\mathbb{R}^2$ into $\mathbb{R}$ and that $\varphi(X,Y)$ is integrable. Then

$$\varphi(E(X),E(Y)) \leq E(\varphi(X,Y))$$
$$(\geq, \text{resp})$$

Proof: Assume that φ is convex:
For all $a,b \geq 0$ with $a+b=1$,

$$\varphi(ax_1+bx_2, ay_1+by_2) = \varphi(a(x_1,y_1)+b(x_2,y_2))$$
$$\leq a\varphi(x_1,y_1) + b\varphi(x_2,y_2).$$

Now let $z(x,y) = \varphi(E(X),E(Y)) + \lambda(x-E(X)) + \mu(y-E(Y))$ be the equation of a plane passing through $(E(X), E(Y), \varphi(E(X),E(Y)))$ and situated below the surface $z = \varphi(x,y)$.

So, $\varphi(E(X),E(Y)) + \lambda(x-E(X)) + \mu(y-E(Y)) \leq \varphi(x,y)$.

$$\Rightarrow \varphi(E(X),E(Y)) + \lambda(X(\omega)-E(X)) + \mu(Y(\omega)-E(Y))$$
$$\leq \varphi(X,Y)(\omega) \quad \forall \omega \in \Omega.$$

If we integrate both sides over $\omega \in \Omega$

$$\varphi(E(X),E(Y))\,\underset{1}{P(\Omega)} + \lambda\overbrace{(E(X) - E(X)P(\Omega))}^{=0}$$
$$+\mu\underbrace{(E(Y) - E(Y)P(\Omega))}_{=0}$$
$$\leq E(\varphi(X,Y))$$

Video 39

$\Rightarrow \varphi(E(X), E(Y)) \le E(\varphi(X,Y))$. $\blacksquare$

Example $\varphi(t) = t^2$, $X : [0,1] \to [0,1]$, $X(t) = t$

$E(X) = \int_0^1 t\,dt = 1/2$, $\varphi(X)(t) = (X(t))^2 = t^2$

$E(\varphi(X)) = \int_0^1 t^2\,dt = 1/3$.

$\varphi(E(X)) = \varphi(1/2) = 1/4 \le 1/3 = E(\varphi(X))$.

Note that $\varphi(t) = t^2$ is a convex function!

Proposition (Markov Inequality)

Let $X \in L_p$, $p \in [1,\infty)$. Then for all $a \in \mathbb{R}$ and $\epsilon > 0$ we have
$$\mathbb{P}(|X-a| \ge \epsilon) \le \frac{1}{\epsilon^p} E(|X-a|^p).$$

In particular, if $a = m(X) = E(X)$ (the mean of X) and $\sigma^2(X) = E((X-m(X))^2)$ (the variance of X), then we get the Chebyshev Inequality
$$\mathbb{P}(|X-m(X)| \ge \epsilon) \le \frac{\sigma^2(X)}{\epsilon^2}.$$

Proof. Let $X \ge 0$ and $X \in L_1$. Then
$$E(X) = \int_\Omega X\,d\mathbb{P} = \int_{\{X \ge \epsilon\}} X\,d\mathbb{P} + \int_{\{X < \epsilon\}} X\,d\mathbb{P}$$
$$\ge \int_{\{X \ge \epsilon\}} \epsilon\,d\mathbb{P} = \epsilon\,\mathbb{P}(X \ge \epsilon).$$

Now replace X by $|X-a|^p$ and note that $|X-a| \geq \epsilon \iff |X-a|^p \geq \epsilon^p$.

$$\mathbb{E}(|X-a|^p) = \int_{\Omega} |X-a|^p \, d\mathbb{P} = \int_{\{|X-a| \geq \epsilon\}} |X-a|^p \, d\mathbb{P} + \int_{\{|X-a| < \epsilon\}} |X-a|^p \, d\mathbb{P}$$

$$\geq \int_{\{|X-a| \geq \epsilon\}} |X-a|^p \, d\mathbb{P}$$

$$\geq \int_{\{|X-a| \geq \epsilon\}} \epsilon^p \, d\mathbb{P}$$

$$= \epsilon^p \, \mathbb{P}(|X-a| \geq \epsilon)$$

So, $\mathbb{P}(|X-a| \geq \epsilon) \leq 1/\epsilon^p \, \mathbb{E}(|X-a|^p)$.

$\blacksquare$

__Proposition__ Let X and Y be independent r.r.v.s on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then the equality
$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ holds in each of the following cases:

i) X and Y are both non-negative (in this case X and Y can be e.r.r.v.s)

ii) $X \in L^p(\Omega, \mathcal{A}, \mathbb{P})$ and $Y \in L^q(\Omega, \mathcal{A}, \mathbb{P})$ with $1/p + 1/q = 1$.

__Definition__ A sequence $\{X_n \mid n \geq 1\} \subseteq L_0(\Omega, \mathcal{A}, \mathbb{P})$ is said to converge almost surely (a.s.) (or almost everywhere a.e.) if, for almost all $\omega \in \Omega$ the sequence $\{X_n(\omega) \mid n \geq 1\}$ converges. In this case, we write
$X_n \xrightarrow{a.s.} X$ if X is the limiting r.r.v.

Definition: A sequence $\{X_n \mid n \geq 1\} \subseteq L_0(\Omega, \mathcal{A}, \mathbb{P})$ is said to converge in probability to a r.v. $X \in L_0(\Omega, \mathcal{A}, \mathbb{P})$, if for all $\epsilon > 0$, $\mathbb{P}(|X_n - X| \geq \epsilon)$ converges to zero. In this case, we write

$$X_n \xrightarrow{pr.} X$$

CHAPTER 6: Moments and Characteristic Functions

§6.1. Moments and Moment Generating Functions:

Definition: i) Let k be a positive integer and X be a r.v. in $L_k(\Omega, \mathcal{A}, \mathbb{P})$ with probability distribution function F_X. Then the k^{th} moment of X or of F_X about $a \in \mathbb{R}$ is defined as $E((X-a)^k)$.

ii) The k^{th} moment about $a=0$, i.e. $E(X^k)$ is called the k^{th} moment of X or of F_X. It is also denoted as m_k or $m_k(X)$.

iii) $E(|X|^k) = (\|X\|_k)^k$ is called the absolute k^{th} moment of X.

iv) The first moment m_1 is also denoted as m.

v) The second central moment about $m \in \mathbb{R}$ $E((X-m)^2)$ is called the variance of X or of F_X and denoted as σ^2 or $\sigma^2(X)$.

Note that

$$\sigma^2(X) = E((X-m)^2)$$
$$= \int_\Omega (X-m)^2 \, d\mathbb{P}$$
$$= \int_\Omega (X^2 - 2mX + m^2) \, d\mathbb{P}$$
$$= \int_\Omega X^2 \, d\mathbb{P} - 2m \int_\Omega X \, d\mathbb{P} + m^2 \int_\Omega d\mathbb{P}$$
$$= m_2 - 2m \cdot m + m^2 \cdot 1$$
$$= m_2 - m^2$$

vi) The positive square root of $\sigma^2 = \sigma^2(X)$ is called the standard deviation of X or of F_X

and is denoted by σ or by $\sigma(X)$ (not to be confused with the σ-algebra generated by X).

vii) The factorial moment of order k is defined by $E(X(X-1)\cdots(X-k+1))$.

Moment Generating Function:

Let X be a r.v. Then the function

$$G_X(u) \doteq E(e^{uX}) = \int_{\mathbb{R}} e^{ux}\, dF_X(x), \quad u \in \mathbb{R},$$

is called the moment generating function of X or F_X.

G_X is a nonnegative function and may take the value $+\infty$. The domain of G_X is defined as $\mathcal{D}(G_X) \doteq \{u \mid G_X(u) < +\infty\}$. It is clear that $0 \in \mathcal{D}(G_X)$.

Proposition: i) $\forall\, a, b \in \mathbb{R}$, $G_{aX+b}(u) = e^{bu}\, G_X(au)$,
ii) If the probability distributions induced by X and $-X$ are the same then G_X is an even function.
iii) G_X is a convex function.

Proof i) $G_{aX+b}(u) = E(e^{u(aX+b)})$
$$= E(e^{ub} e^{uaX})$$
$$= e^{ub} E(e^{uaX})$$
$$= e^{ub} G_X(au)$$

ii) If $F_X = F_{-X}$ then

Video 40

$$G_X(-u) = E(e^{-uX}) = E(e^{u(-X)}) = G_{-X}(u)$$

$$\Rightarrow G_X(-u) = G_{-X}(u) = \int_{\mathbb{R}} e^{ux}\, dF_{-X}(x)$$
$$= \int_{\mathbb{R}} e^{ux}\, dF_X(x)$$
$$= G_X(u).$$

iii) Let $a, b \in \mathbb{R}_+$ s.t. $a+b=1$. Then since e^{t} is a convex function we have

$$G_X(au_1 + bu_2) = E\left(e^{(au_1+bu_2)X}\right)$$
$$\leq E\left(a e^{u_1 X} + b e^{u_2 X}\right)$$
$$= a E(e^{u_1 X}) + b E(e^{u_2 X})$$
$$= a G_X(u_1) + b G_X(u_2).$$ ∎

Theorem: Two r.v.s have the same moment generating function if and only if they have the same distribution function.

Proposition: Let X and Y be two independent r.v.s. Then the moment generating function of $X+Y$ is given by
$$G_{X+Y}(u) = G_X(u) \cdot G_Y(u).$$

Proof: $G_{X+Y}(u) = E(e^{u(X+Y)}) = E(e^{uX} \cdot e^{uY})$
$$= E(e^{uX}) \cdot E(e^{uY})$$
because since X and Y are independent so are e^{uX} and e^{uY}. ∎

Proposition: Let X be a r.v. with moment generating function G_X. Assume that there is a negative number u_1 and a positive number u_2 so that $(u_1, u_2) \subseteq \mathcal{D}(G_X)$. Then

i) $\forall k \geq 1$, $E(|X|^k) < \infty$.

ii) $\forall u \in (-s, s)$, $0 < s < s_0 = \min(-u_1, u_2)$, we have

$$G_X(u) = 1 + E(X)\frac{u}{1!} + E(X^2)\frac{u^2}{2!} + \dots + E(X^n)\frac{u^n}{n!} + \dots$$

iii) $\forall k \geq 1$, $G_X^{(k)}(0) = E(X^k)$.

Proof: Let's denote F_X by F for simplicity.

i) Note that $\dfrac{u^k |x|^k}{k!} \leqslant e^{u|x|}$ and thus $u^k |x|^k \leq k!\, e^{u|x|}$ for all $k \geq 1$, $u > 0$.

Hence,
$$E(|X|^k) \leqslant \frac{k!}{u^k} \int_{-\infty}^{\infty} e^{u|x|}\, dF(x).$$

$$= \frac{k!}{u^k}\left(\int_0^{\infty} e^{ux}\, dF(x) + \int_{-\infty}^{0} e^{-ux}\, dF(x)\right)$$

$$< \frac{k!}{u^k}\left(\int_{-\infty}^{\infty} e^{ux}\, dF(x) + \int_{-\infty}^{0} e^{-ux}\, dF(x) - \int_{-\infty}^{0} e^{ux}\, dF(x)\right)$$

$$= \frac{k!}{u^k}\left(\int_{-\infty}^{\infty} e^{ux}\, dF(x) + \int_{-\infty}^{0} e^{-ux}\, dF(x) + \int_{0}^{-\infty} e^{ux}\, dF(x)\right)$$

$$= \frac{k!}{u^k}\left(\int_{-\infty}^{\infty} e^{ux}\, dF(x) + \int_{-\infty}^{0} e^{-ux}\, dF(x) + \int_{0}^{\infty} e^{-ux}\, dF(x)\right)$$

$$= \frac{k!}{u^k}\left(G_X(u) + G_X(-u)\right)$$

The last expression is finite for

$u \in (-s,s) \subseteq (u_1, u_2) \subseteq \mathcal{D}(G_X)$. This finishes the proof. ■

ii) We have $G_X(u) = \int_{-\infty}^{+\infty} e^{ux}\, dF(x) = \int_{-\infty}^{\infty} \left(\sum_{k=0}^{\infty} \frac{(ux)^k}{k!} \right) dF(x)$

We need to show that the integral and the summation signs in the last term can be interchanged. To see this let

$$h(x) = e^{ux} \quad \text{and} \quad h_n(x) = \sum_{k=0}^{n} (ux)^k / k! \ , \ \forall n \geq 1.$$

We know that $\forall x \in \mathbb{R}$, $\lim_{n \to \infty} h_n(x) = h(x)$ and

hence, $|h_n(x)| \leq \sum_{k=0}^{n} \frac{|ux|^k}{k!} \leq \sum_{k=0}^{\infty} \frac{|ux|^k}{k!} = e^{|ux|} \leq e^{s|x|}$.

Now by the Dominated Convergence Theorem we get

$$G_X(u) = \sum_{k=0}^{\infty} \int_{-\infty}^{\infty} \frac{u^k x^k}{k!}\, dF(x) = \sum_{k=0}^{\infty} \frac{u^k}{k!} E(X^k).$$

iii) The series of ii) can be differentiated term by term and this gives

$$G_X^{(k)}(u) = \sum_{n \geq k} n(n-1)\cdots(n-k+1) E(X^n) \frac{u^{n-k}}{n!}.$$

Then, $G_X^{(k)}(0) = \underbrace{k(k-1)\cdots(k-k+1)}_{= k!} E(X^k) \frac{u^0}{k!}$

$= E(X^k)$. ■

Video 41

Generating Functions:

Definition: Let X be a r.v. taking values in $\mathbb{N}$. Let $p_k = \mathbb{P}(\{X = k\})$. Then

$$G_X(u) = \int_{\mathbb{R}} e^{ux}\, dF_X(x) = \sum_{k=0}^{\infty} p_k e^{ku} = \sum_{k=0}^{\infty} p_k < +\infty$$

Define the function $g_X(z) = \sum_{k=0}^{\infty} p_k z^k$, $z \in \mathbb{C}$,

whenever the series is convergent. It will be called the probability generating function of X.

Proposition: $g_X(z)$ is absolutely convergent on the closed unit disc $\{z \in \mathbb{C} \mid |z| \leq 1\}$. Therefore, it has analytic derivatives on the open unit disc $\{z \in \mathbb{C} \mid |z| < 1\}$. Also, the n^{th} derivative is $g_X^{(n)}(z) = \sum_{k \geq n} k(k-1)\cdots(k-n+1) z^{k-n}$, which gives

$$g_X^{(n)}(0) = n!\, p_n, \quad \forall n \in \mathbb{N}.$$

Proof Since $G_X(0) = \sum_{k=0}^{\infty} p_k = 1 < +\infty$ then

the series $g_X(z) = \sum_{k=0}^{\infty} p_k z^k$ is convergent for

all $|z| \leq 1$. In particular, $G_X(z)$ is analytic on $|z| < 1$. ∎

Lemma (Abel's Theorem)

1) Assume that we define $f(x)$ on $(-1,1)$ by the real power series $\sum_{n \geq 0} a_n x^n$ supposed to

converge on $(-1,1)$. If the series also converges at $x=1$, then the limit $f(1-0)=\lim_{x\to 1^-} f(x)$ exists and equals to $\sum_{n=0}^{\infty} a_n$.

2) Assume that for all $n\in\mathbb{N}$, $a_n\geq 0$ and $\sum_{n=0}^{\infty} a_n<\infty$ and let $f(x)$ denote the series $\sum_{n=0}^{\infty} a_n x^n$, which converges for all $x\in(-1,1)$. Then $f(1-0)$ exists and equals to $\sum_{n=0}^{\infty} a_n$.

Proof: (1) is known as Abel's Theorem. Part (2) follows from (1). ∎

Proposition Let $g_X(u)$ be the generating function defined above for $u\in(-1,1)$. For $n\geq 1$, $g_X^{(n)}(1-0)$ exists if and only if

$E(X(X-1)\ldots(X-n+1))<\infty$, in other words, if the factorial moment of order n exists. In this case, $g_X^{(n)}(1-0)=E(X(X-1)\ldots(X-n+1))$.

Proof. By the proposition before the Abel's Thm $g_X^{(n)}(1-0)$ exists. So by Abel's Theorem this limit is equal to

$$\sum_{k\geq n} k(k-1)\ldots(k-n+1)\,p_k=E(X(X-1)\ldots(X-n+1))<+\infty.$$

Conversely, if the expectation is finite, i.e., if the factorial moment of order n exists, then by part (2) of the above lemma $g_X(1-0)$ exists and equals $E(X(X-1)\ldots(X-n+1))$. ∎

§6.2. Laplace Transform This is the analog of moment generating function for non negative r.v.s.

Definition: Let X be a nonnegative rrv with distribution function $F_X(x)$, so that $F_X(x)=0$ for $x\in(-\infty,0)$. The Laplace transform of F_X is defined by

$$\mathcal{L}_X(s) \doteq \mathbb{E}(e^{-sX}) = \int_{\mathbb{R}} e^{-sx}\, dF_X(x), \quad s\in\mathbb{C}, \operatorname{Re} s > 0.$$

Proposition: i) For $a>0$ and $b\geq 0$ we have

$$\mathcal{L}_{aX+b}(s) = e^{-bs}\mathcal{L}_X(as) = e^{-bs}\int_{\mathbb{R}^+} e^{-su}\, dF_X\left(\frac{u}{a}\right),$$ where

$F_X\left(\frac{u}{a}\right)$ is the distribution function of aX.

ii) Let X and Y be independent nonnegative rrvs. Then $\mathcal{L}_{X+Y}(s) = \mathcal{L}_X(s)\,\mathcal{L}_Y(s)$, for all $s\in R=\{s\in\mathbb{C} \mid \operatorname{Re} s>0\}$.

In other words, $\mathcal{L}_{X+Y}$ is the Laplace transform of

$$F(z) = \int_{[0,z]} F_X(z-y)\, dF_Y(y) = \int_{[0,z]} F_Y(z-x)\, dF_X(x).$$

iii) $\mathcal{L}(s)$ is analytic on R. Its n^{th} derivative satisfies

a) $$\mathcal{L}^{(n)}(s) = \int_{\mathbb{R}^+} (-x)^n e^{-sx}\, dF(x), \quad s\in\mathbb{R}.$$

b) $$\mathcal{L}^{(n)}(0+) = \lim_{s\to 0^+} \mathcal{L}^{(n)}(s), \quad s\in(0,\infty)$$ exists and is finite

if and only if the n^{th} moment exists and in this case

$$\mathcal{L}^{(n)}(0) = \mathcal{L}^{(n)}(0+) = \int_{\mathbb{R}_+} (-x)^n\, dF(x) = (-1)^n m_n$$

c) Assume that the n^{th} moment of F exists. Then $\mathcal{L}$ admits the following Taylor expansion

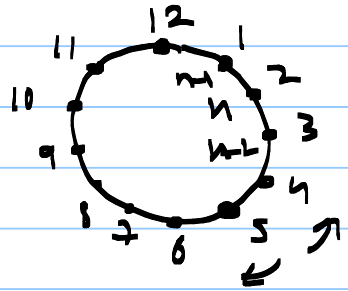
$$\mathcal{L}(s) = \sum_{k=0}^{n} (-1)^k \frac{s^k}{k!} \mathbb{E}(X^k) + o(s), \text{ for small } s>0.$$

Video 42

Example: 1) (Stochastic Process)

Alice has 7 marbles and Bob has 5 marbles. They play the following game. They toss a coin and whoever wins gets one marble from the other person. The first person who gets all the 12 marbles wins the game. What is the average number of tossings for a game to finish?

Solution:



Think of the following experiment: Each second a particle comes up from the hole 5 and it moves each second one step in either direction with equal probability. A particle that reaches 12 disappears.

For any $k \in \{1,2,\dots,12\} = \Omega$ and $n \geq 0$ any integer let

$E_k(n)$ = "the probability for the particle of being at the kth place at the end of the nth second (assuming it is at 5th place at $t=0$)."

$$E_1(n) = \frac{1}{2} E_2(n-1)$$

$$E_2(n) = \frac{1}{2} E_1(n-1) + \frac{1}{2} E_3(n-1)$$

$$\vdots$$

$$E_{10}(n) = \frac{1}{2} E_9(n-1) + \frac{1}{2} E_{11}(n-1)$$

$$E_{11}(n) = \frac{1}{2} E_{10}(n-1)$$

$$E_{12}(n) = \frac{1}{2} E_1(n-1) + \frac{1}{2} E_{11}(n-1).$$

Let
$$E(n) = \begin{bmatrix} E_1(n) \\ \vdots \\ E_{12}(n) \end{bmatrix}$$

$$E(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then we can the matrix equation $\mathcal{E}_n = A\mathcal{E}_{n-1}$, where

$$A = \begin{bmatrix} 0 & 1/2 & 0 & \cdots & \cdots & 0 & 0 \\ 1/2 & 0 & 1/2 & \cdots & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 1/2 & \vdots \\ & & & \ddots & \ddots & 0 & \\ 1/2 & 0 & 0 & \cdots & \cdots & 1/2 & 0 \end{bmatrix}$$

$$\mathcal{E}_n = A\mathcal{E}_{n-1} = A(A\,\mathcal{E}_{n-2}) = A^2\mathcal{E}_{n-2} = \cdots = A^n\,\mathcal{E}(0)$$

So, for the particle who comes up from 5 at $t=0$, its probability for being at the k^{th} place at the end of the n^{th} second is $\mathcal{E}_k(n)$ where $\mathcal{E}(n) = A^n\,\mathcal{E}(0)$.

In particular, the average number of particles at the k^{th} place at the end of the n^{th} second is the sum

$$\underline{\underline{\mathcal{E}_k(0)}} + \cdots\cdots + \mathcal{E}_k(n-1) + \mathcal{E}_k(n).$$

Hence, the limiting number of particles at the k^{th} place is $\displaystyle\sum_{i=0}^{\infty} \mathcal{E}_k(i) = \Big(\sum_{i=0}^{\infty} A^i\,\mathcal{E}(0)\Big)_k \leftarrow$ k^{th} entry of the column $\big(\sum_{i=0}^{\infty} A^i\,\mathcal{E}(0)\big)$

$$= \Big(\Big(\sum_{i=0}^{\infty} A^i\Big)\,\mathcal{E}(0)\Big)_k$$

The matrix A has norm $\|A\| < 1$ and thus $\displaystyle\sum_{i=0}^{\infty} A^i$ is convergent and equals $\dfrac{1}{1-A} = (Id - A)^{-1}!$

$$
(I_d - A)^{-1} = \begin{bmatrix} 7/6 \\ 7/3 \\ 7/2 \\ 14/3 \\ 35/6 \\ 5 \\ 25/6 \\ 10/3 \\ 5/2 \\ 5/3 \\ 5/6 \\ 1 \end{bmatrix}
$$

5th column.

$$
E_T(\infty) \doteq \left(\sum_{i=0}^{\infty} A^i\right) E(0) = (I_d - A)^{-1} E(0) = \text{the 5th column of } (I_d - A)^{-1}.
$$

$$
E_T(\infty) = \begin{bmatrix} E_1(\infty) \\ \vdots \\ E_{12}(\infty) \end{bmatrix}.
$$

So the total number of particles on the experiment table is $E_1(\infty) + \dots + E_{12}(\infty) = 35$.

Hence the average life span of a particle is 35 seconds. Thus, the game Alice and Bob are playing will end approximately at the end of the 35th tossing.