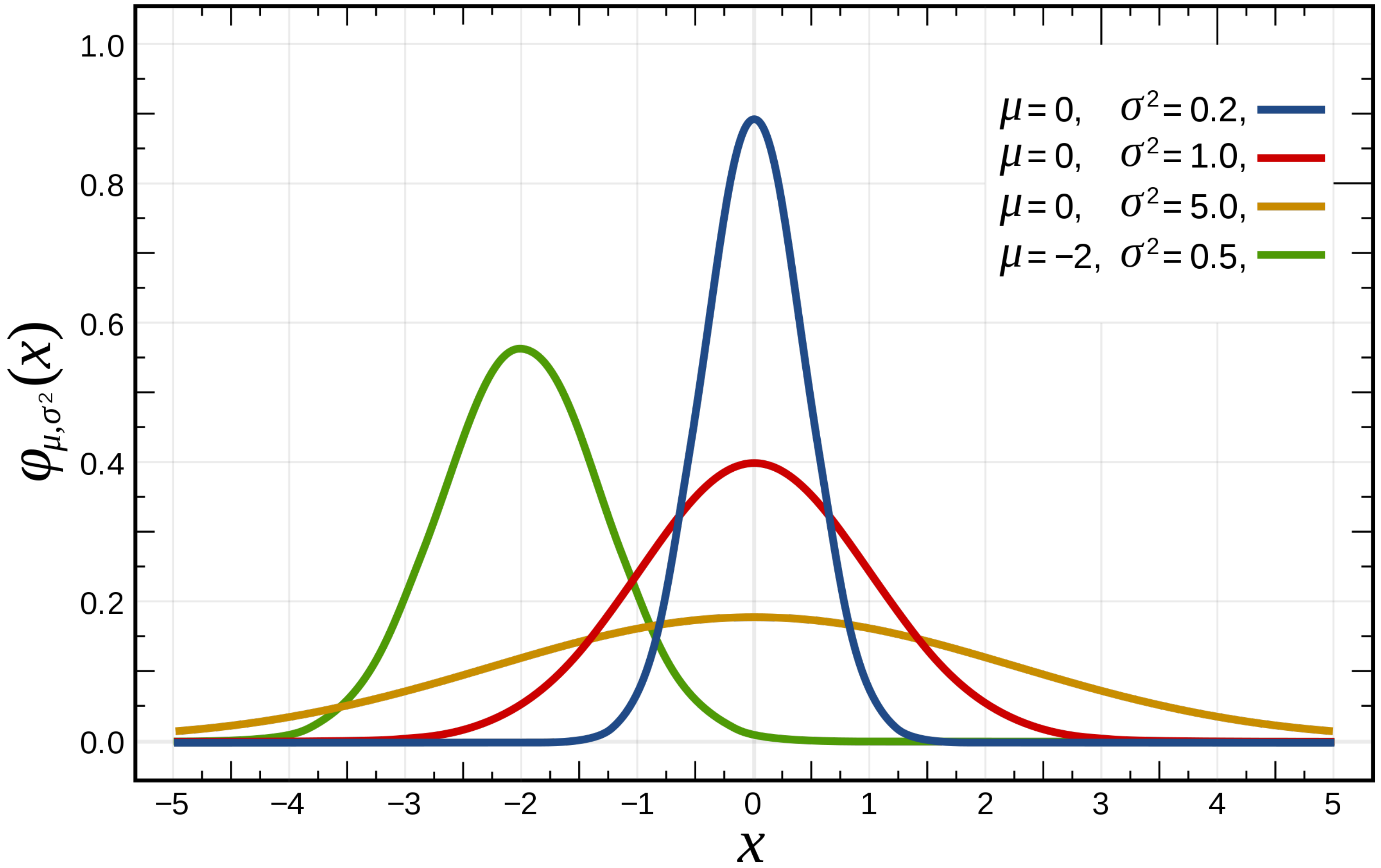

Introduction to Probability Theory

Math 531

Author: Chris Cai/ Linrong Cai

Institute: University of Wisconisn Madison

Date: May 15, 2023

Version: 1.0

Instructor: Benedek Valkó

The theory of probabilities is at bottom nothing but common sense reduced to calculus; it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which of times they are unable to account.

Contents

Chapter 1 Events and Probability

1.1 Kolmogorov's axioms and probability space

These are Kolmogorov's axioms for probability theory.

Definition 1.1

A ***probability space*** *is a triple* $(\Omega, \mathcal{F}, P)$*, with the following components.*
(a) Ω *is a set, called the* ***sample space****.*
(b) $\mathcal{F}$ *is a collection of subsets of* Ω*. Members of* $\mathcal{F}$ *are called events.* $\mathcal{F}$ *is assumed to be a* σ*-algebra, which means that it satisfies the following three properties.*
(b.1) $\Omega \in \mathcal{F}$*. That is, the whole sample space itself is an event.*
(b.2) If $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$*.*
(b.3) If $\{A_k\}_{1\le k<\infty}$ *is a sequence of members of* $\mathcal{F}$*, then their union* $\bigcup_{k=1}^{\infty} A_k$ *is also a member of* $\mathcal{F}$*.*
(c) P *is a function from* $\mathcal{F}$ *into real numbers, called the* ***probability measure*** *.* P *satisfies the following axioms.*
(c.1) $0 \le P(A) \le 1$ *for each event* $A \in \mathcal{F}$*.*
(c.2) $P(\varnothing) = 0$ *and* $P(\Omega) = 1$*.*
(c.3) If $\{A_k\}_{1\le k<\infty}$ *is a sequence of pairwise disjoint events then*

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P\left(A_k\right).$$

Note *In mathematical analysis and in probability theory, a* $\boldsymbol{\sigma}$*-algebra (also* $\boldsymbol{\sigma}$*-field) on a set* X *is a nonempty collection* Σ *of subsets of* X *closed under complement, countable unions, and countable intersections. The ordered pair* (X, Σ) *is called a measurable space.*

1.2 Inclusion Exclusion Principle

Theorem 1.1 (inclusion-exclusion principle)

Let $A_1, A_2, A_3, \ldots$ *be events in some probability space* $(\Omega, \mathcal{F}, P)$*. Then for each integer* $n \ge 2$*,*

$$\begin{aligned} P\left(A_1 \cup \cdots \cup A_n\right) = & \sum_{i=1}^{n} P\left(A_i\right) - \sum_{1\le i_1<i_2\le n} P\left(A_{i_1} \cap A_{i_2}\right) \\ & + \sum_{1\le i_1<i_2<i_3\le n} P\left(A_{i_1} \cap A_{i_2} \cap A_{i_3}\right) \\ & - \sum_{1\le i_1<i_2<i_3<i_4\le n} P\left(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_{i_4}\right) \\ & + \cdots + (-1)^{n+1} P\left(A_1 \cap \cdots \cap A_n\right). \\ = & \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\le i_1<\cdots<i_k\le n} P\left(A_{i_1} \cap \cdots \cap A_{i_k}\right). \end{aligned}$$

This is called the inclusion-exclusion identity.

1.3 Monotonicity and Countable Subadditivity

Proposition 1.1

Let $A, B, A_1, A_2, A_3, \ldots$ be events in some probability space $(\Omega, \mathcal{F}, P)$
(i) Monotonicity: if $A \subset B$ then $P(A) \leq P(B)$.
(ii) Countable subadditivity: for any sequence of events $\{A_k\}$,

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P\left(A_k\right).$$

Countable subadditivity generalizes the countable additivity axiom in a natural way. Its truth should be fairly obvious because the union $\bigcup_{k=1}^{\infty} A_k$ can have overlaps whose probabilities are then counted several times over in the sum $\sum_{i=1}^{\infty} P\left(A_k\right)$. By taking $A_k = \varnothing$ for all $k > n$ we get a finite version of subadditivity:

$$P\left(A_1 \cup \cdots \cup A_n\right) \leq P\left(A_1\right) + \cdots + P\left(A_n\right)$$

valid for all events $A_1, \ldots, A_n$. ♣

Corollary 1.1

Let $\{A_k\}$ be a sequence of events on $(\Omega, \mathcal{F}, P)$.
(i) If $P\left(A_k\right) = 0$ for all k, then $P\left(\bigcup_k A_k\right) = 0$.
(ii) If $P\left(A_k\right) = 1$ for all k, then $P\left(\bigcap_k A_k\right) = 1$. ♡

1.4 Continuity of Probability

Definition 1.2

Suppose $\{A_k\}_{k \in \mathbb{Z}_{>0}}, \{B_k\}_{k \in \mathbb{Z}_{>0}}, A$, and B are events in a probability space $(\Omega, \mathcal{F}, P)$. We say that A_k increases up to A and use the notation

$$A_k \nearrow A$$

if the events A_k are nested nondecreasing, which means that $A_1 \subset A_2 \subset A_3 \subset \cdots \subset A_k \subset \cdots$, and $A = \bigcup_k A_k$. Figure 1 illustrates. Analogously, we say that B_k decreases down to B and use the notation

$$B_k \searrow B$$

if the events B_k are nested nonincreasing, which means that $B_1 \supset B_2 \supset B_3 \supset \cdots \supset B_k \supset \cdots$, and $B = \bigcap_k B_k$ ♠

Theorem 1.2

If $A_k \nearrow A$ or $A_k \searrow A$, then the probabilities converge: $\lim_{k \rightarrow \infty} P\left(A_k\right) = P(A)$ ♡

1.5 Conditional Probability

Definition 1.3 (conditional probability)

Let B be an event on the probability space $(\Omega, \mathcal{F}, P)$ such that $P(B) > 0$. Then for all events $A \in \mathcal{F}$ the conditional probability of A given B is defined as

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

Proposition 1.2

Let B be an event on the probability space $(\Omega, \mathcal{F}, P)$ such that $P(B) > 0$. Then as a function of the event A, $P(A \mid B)$ is a probability measure on $(\Omega, \mathcal{F})$

Theorem 1.3

In each statement below all events are on the same probability space$(\Omega, \mathcal{F}, P)$

(a) Let A and B be two events and assume $P(B) > 0$. Then

$$P(AB) = P(B)P(A \mid B)$$

Let $A_1, \ldots, A_n$ be events and assume $P\left(A_1 \cdots A_{n-1}\right) > 0$. Then

$$P\left(A_1 A_2 \cdots A_n\right) = P\left(A_1\right) P\left(A_2 \mid A_1\right) P\left(A_3 \mid A_1 A_2\right) \cdots P\left(A_n \mid A_1 \cdots A_{n-1}\right)$$

(b) Let $\{B_i\}$ be a countable partition of Ω and A any event. Then

$$P(A) = \sum_{i : P(B_i) > 0} P\left(A \mid B_i\right) P\left(B_i\right).$$

The sum above ranges over those indices i such that $P\left(B_i\right) > 0$.

1.6 Bayes' Formula

Theorem 1.4 (Bayes' formula)

Let $\{B_k\}$ be a countable partition of the sample space Ω. Then for any event A with $P(A) > 0$ and each k such that $P\left(B_k\right) > 0$,

$$P\left(B_k \mid A\right) = \frac{P\left(AB_k\right)}{P(A)} = \frac{P\left(A \mid B_k\right) P\left(B_k\right)}{\sum_{i : P(B_i) > 0} P\left(A \mid B_i\right) P\left(B_i\right)}$$

1.7 Independent Events

Definition 1.4

Two events A and B are independent if

$$P(AB) = P(A)P(B)$$

Theorem 1.5

Suppose that A and B are independent events. Then the same is true for each of these pairs: A^c and B, A and B^c, and A^c and B^c.

The definition of independence of more than two events requires that the product property hold for any subcollection of events.

Definition 1.5

(a) Events $A_1, \ldots, A_n$ are independent (or mutually independent) if for every collection $A_{i_1}, \ldots, A_{i_k}$, where $2 \leq k \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$

$$P\left(A_{i_1} A_{i_2} \cdots A_{i_k}\right) = P\left(A_{i_1}\right) P\left(A_{i_2}\right) \cdots P\left(A_{i_k}\right)$$

(b) Let $\{A_k\}_{k \in \mathbb{Z}_{>0}}$ be an infinite sequence of events in a probability space $(\Omega, \mathcal{F}, P)$. Then events $\{A_k\}_{k \in \mathbb{Z}_{>0}}$ are independent if for each $n \in \mathbb{Z}_{>0}$, events $A_1, \ldots, A_n$ are independent.

Theorem 1.6

(a) Suppose events $A_1, \ldots, A_n$ are mutually independent. Then for every collection $A_{i_1}, \ldots, A_{i_k}$, where $2 \leq k \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, we have

$$P\left(A_{i_1}^* A_{i_2}^* \cdots A_{i_k}^*\right) = P\left(A_{i_1}^*\right) P\left(A_{i_2}^*\right) \cdots P\left(A_{i_k}^*\right)$$

where each A_i^* can represent either A_i or A_i^c.
(b) Let $\{A_k\}_{k \geq 1}$ be a finite or infinite sequence of independent events. Let $0 = k_0 < k_1 < \cdots < k_n$ be integers. Let $B_1, \ldots, B_n$ be events constructed from the A_ks so that, for each $j = 1, \ldots, n$, B_j is formed by applying set operations to $A_{k_{j-1}+1}, \ldots, A_{k_j}$. Then the events $B_1, \ldots, B_n$ are independent.

Definition 1.6

Let $A_1, \ldots, A_n$ and B be events on $(\Omega, \mathcal{F}, P)$ and assume $P(B) > 0$. Then events $A_1, \ldots, A_n$ are conditionally independent, given B, if for every collection $A_{i_1}, \ldots, A_{i_k}$, where $2 \leq k \leq n$ and $1 \leq i_1 < i_2 < \cdots < i_k \leq n$,

$$P\left(A_{i_1} A_{i_2} \cdots A_{i_k} \mid B\right) = \prod_{j=1}^{k} P\left(A_{i_j} \mid B\right)$$

Chapter 2 Random Variables and Probability Distributions

2.1 Random Variables

Often we are interested in some numerical value associated to the outcome of a random experiment. This just means that we are interested in the value of a function that maps the elements of the sample space into the real numbers. These functions are called random variables.

Definition 2.1

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable on Ω is a real valued function $X : \Omega \to \mathbb{R}$, for which $\{\omega \in \Omega : X(\omega) \leq c\} \in \mathcal{F}$ for any $c \in \mathbb{R}$.

There is a simple way to encode an events as a random variable.

Definition 2.2

Let B be an event in a probability space. Then the indicator function (or indicator random variable) of B is defined as the function

$$I_B(\omega) = \begin{cases} 1, & \text{if } \omega \in B \\ 0, & \text{if } \omega \notin B \end{cases}$$

Note *Note that I_B is a random variable.*

2.2 Probability Distributions

Through the probabilities of events of type $\{X \in B\}$, a random variable induces a probability measure on the real line.

Definition 2.3

Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. The probability distribution of X is the probability measure μ on $\mathbb{R}$ defined by

$$\mu(B) = P(X \in B)$$

for Borel subsets B of $\mathbb{R}$.

Note *In mathematics, a Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. Borel sets are named after Émile Borel.*