Introduction to Measure Theory

Sangdon Park

POSTECH

February 25, 2025

Motivation

1. σ-algebra? distribution? induced distribution? measurable space v.s. probability space?
2. Are we using statistical terms correctly?
3. Why is this definition valid?

Definition

A random variable X is said to have a Binomial(n, p) distribution if

$$\mathbb{P}(X = m) := \binom{n}{m} p^m (1-p)^{n-m}.$$

4. Rigorus proofs

Measure?

The meaning of “measure” in “measure theory” is the same as the “measure” in the following:

- How to *measure* the height of a boy?
- How to *measure* the legnth of the width of a table?
- How to *measure* the size of an area?
- How to *measure* the size of a discrete set?

We will learn the rigorous definition of “measure”.

Algebra

Definition (Algebra)

Let Ω be a nonempty set. A set $\mathcal{F}$ is an **algebra** of sets on Ω if it is a nonempty collection of subsets of a set Ω that satisfies

1. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (*i.e.*, $\mathcal{F}$ is closed under complements), and
2. if $A_i, \ldots, A_n \in \mathcal{F}$, then $\cup_{i=1}^n A_i \in \mathcal{F}$ (*i.e.*, $\mathcal{F}$ is closed under finite unions).

- Mathematicians decided to call *a set of sets with two properties* an **algebra**.

σ-algebra

Definition (σ-algebra)

Let Ω be a nonempty set. A set $\mathcal{F}$ is a σ-**algebra** of sets on Ω if it is a nonempty collection of subsets of a set Ω that satisfies

1. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (*i.e.*, $\mathcal{F}$ is closed under complements), and
2. if $A_i \in \mathcal{F}$ is a countable sequence of sets, then $\cup_i A_i \in \mathcal{F}$ (*i.e.*, $\mathcal{F}$ is closed under countable unions).

- These implies that a σ-algebra is closed under countable intersections (*i.e.*, $A_i \in \mathcal{F} \implies \cap_i A_i = (\cup_i A_i^c)^c \in \mathcal{F}$).
- The above two properties are minimum requirements to define a “measure” (*e.g.*, a ruler).

Measurable Space

Definition

A tuple $(\Omega, \mathcal{F})$ is a **measurable space** if Ω is a non-empty set and $\mathcal{F}$ is a σ-algebra.

- A measurable space is a space on which we can put a “measure”.
  - A σ-algebra $\mathcal{F}$ is a good enough set to put a “measure”
  - It is not a “measure” space – a “measure” is not yet defined.

Wait! Why Do We Need These Complicatd Definitions?

- A *non-measurable set* is a set which cannot be asigned a meaningful “volume”.
- There exists a non-measurable subset of $\mathbb{R}$ in Zermelo–Fraenkel set theory.
- σ-algebra is suffficiently huge collection to define a measure.

Measure

Definition

A **measure** μ on a measureble space $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to \mathbb{R}_{\geq 0}$ where

1. $\mu(\emptyset) = 0$ and
2. if $A_i \in \mathcal{F}$ is a countable sequence of disjoint sets, then

$$\mu(\cup_i A_i) = \sum_i \mu(A_i).$$

- If μ is a measure on a measurable space $(\Omega, \mathcal{F})$, then $(\Omega, \mathcal{F}, \mu)$ is a **measure space**.
- If $\mu(\Omega) = 1$, we call μ a **probability measure**, denoted by $\mathbb{P}$.

Probability Space

Definition

A **probability space** is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ with a probability measure $\mathbb{P}$, where

- Ω is a set of "outcomes",
- $\mathcal{F}$ is a set of "events", and
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a function that assigns probabilities to events.

Properties of A Measure

The properties of a measure is derived from the definition of the measure.

Theorem

Let $\mu : \mathcal{F} \to \mathbb{R}$ *be a measure on* $(\Omega, \mathcal{F})$.

1. *(Monotonicity) If* $A \subseteq B$, *then* $\mu(A) \leq \mu(B)$.
2. *(Subadditivity) If* $A \subseteq \cup_{m=1}^{\infty} A_m$, *then* $\mu(A) \leq \sum_{m=1}^{\infty} \mu(A_m)$.
3. ...

Properties of A Measure

The properties of a measure is derived from the definition of the measure.

Theorem

Let $\mu : \mathcal{F} \to \mathbb{R}$ be a measure on $(\Omega, \mathcal{F})$.

1. *(Monotonicity) If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.*
2. *(Subadditivity) If $A \subseteq \cup_{m=1}^{\infty} A_m$, then $\mu(A) \leq \sum_{m=1}^{\infty} \mu(A_m)$.*
3. ...

Proof (Monotonicity).

Let $B - A = B \cap A^c$ be the difference of the two sets. Using $+$ to denote disjoint union, $B = A + (B - A)$ so

$$\mu(B) = \mu(A) + \mu(B - A) \geq \mu(A)$$

due to the definition of a measure. □

Measure On The Real Line

How to design a measure? A measure function defines a measure.

Definition

A **Stieltjes measure function** is a function $F : \mathbb{R} \to \mathbb{R}$ where

1. F is non-decreasing and
2. F is right-continuous, *i.e.*,

$$\lim_{y \downarrow x} F(y) = F(x).$$

Thomas Joannes Stieltjes
(known for Riemann–Stieltjes integral)

Measure On The Real Line

Can we define a measure by using the Stieltjes measure function?

Theorem

Given a Stieltjes measure function F, there is a unique measure μ on $(\mathbb{R}, \mathcal{R})$ with

$$\mu((a, b]) = F(b) - F(a).$$

- When $F(x) = x$, the resulting measure is called **Lebesgue measure**.
- *e.g.*, a length of an interval is a measure.

Random Variables

Definition (measurable map)

A function $X : \Omega \to S$ is a **measurable map** from a measurable space $(\Omega, \mathcal{F})$ to a measurable space $(S, \mathcal{S})$ if

$$X^{-1}(B) := \{\omega \in \Omega \mid X(w) \in B\} \in \mathcal{F} \text{ for all } B \in \mathcal{S}.$$

- Connect two measurable spaces – don't need to define measures again.
- Help to *reuse* a measure defined on the measurable space $(\Omega, \mathcal{F})$.
  - The measure on the new space is well-defined based on the measure on the old space.
- When $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$,
  - if $d > 1$, then X is called a **random vector** and
  - if $d = 1$, then X is called a **random variable**.
- If Ω is a discrete probability space, then any function $X : \Omega \to \mathbb{R}$ is a random variable.

Distribution

Definition (distribution)

An induced probability measure μ on $(\mathbb{R}, \mathcal{R})$ by a random variable $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{R})$ is called a **distribution**, *i.e.*, for any $B \in \mathcal{R}$

$$\mu(B) := \mathbb{P}(X^{-1}(B)).$$

- Redefine a measure over an easy space (*i.e.*, $\mathbb{R}$) and call it a "distribution".
  - A distribution is a measure.
- A distribution depends on an random variable.
- Is μ a probability measure? Let's only check the second property of a measure. For any disjoint sets B_i,

$$\mu(\cup_i B_i) = \mathbb{P}(X^{-1}(\cup_i B_i)) = \mathbb{P}(\cup_i X^{-1}(B_i)) = \sum_i \mathbb{P}(X^{-1}(B_i)) = \sum_i \mu(B_i).$$

- How to define a probability measure? One way is...
  - Define a simple "distribution function" over $\mathbb{R}$ (*e.g.*, the Gaussian distribution).

Distribution Functions

Definition

A (usual) **distribution function** of a random variable $X : \mathbb{R} \to \mathbb{R}$ is the function $F : \mathbb{R} \to \mathbb{R}$ defined by

$$F(x) := \mathbb{P}(X \leq x).$$

- a.k.a. a cumulative distribution function (CDF)
- In the real line, due to the monotonicity of a measure, CDF is enough to define a measure.

Density Functions

Definition

X has a **density function** f_X if a distribution function $F(x) = \mathbb{P}(X \leq x)$ has the form

$$F(x) = \int_{-\infty}^{x} f_X(y)\mathrm{d}y.$$

- Normal distribution: $f_X(x) := \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Once a density function is defined, the probability measure is indirectly defined.
  - We don't need to define the probability measure in the original space directly (thanks to σ-algebra, measure, measurable map, ...)

More Facts on Random Variables

Theorem

If $X_1, \ldots, X_n$ *are random variables and* $f : (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$ *is measurable, then* $f(X_1, \ldots, X_n)$ *is a random variable.*

More Facts on Random Variables

Theorem

If $X_1, \ldots, X_n$ *are random variables and* $f : (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$ *is measurable, then* $f(X_1, \ldots, X_n)$ *is a random variable.*

Theorem

If $X_1, \ldots, X_n$ *are random variables, then* $X_1 + \cdots + X_n$ *is a random variable.*

More Facts on Random Variables

Theorem

If $X_1, \ldots, X_n$ *are random variables and* $f : (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$ *is measurable, then* $f(X_1, \ldots, X_n)$ *is a random variable.*

Theorem

If $X_1, \ldots, X_n$ *are random variables, then* $X_1 + \cdots + X_n$ *is a random variable.*

Theorem (product measure)

If $(\Omega_i, \mathcal{F}_i, \mu_i)$ *for* $i = 1, \ldots, n$ *are measure spaces and* $\Omega := \Omega_1 \times \ldots \Omega_n$*, there is a unique measure* μ *on* $(\prod_i \Omega_i, \prod_i \mathcal{F}_i)$ *where*

$$\mu(A_1 \times \cdots \times A_n) = \prod_i \mu_i(A_i)$$

for any $A_i \in \mathcal{F}_i$*.*

Use Case: Binomial Distribution

Definition

A random variable X is said to have a Binomial(n, p) distribution if

$$\mathbb{P}(X = m) = f_X(m) := \binom{n}{m} p^m (1-p)^{n-m}.$$

- The Binomial random variable is a sum of Bernoulli random variables.
- It is usually explained via a sequence of coin flipping.
- We define a probability measure on the original space.
- How can it be redefined over $\mathbb{N}$?
- Why do we have this Binomial distribution?
  - We will interpret this in a measure-theoretic perspective.

Use Case: Binomial Distribution I

Proof Sketch:

1. We have a probability space $(\Omega, \mathcal{F}, \mathbb{P}_0)$.
   - $\Omega := \{\text{"S"}, \text{"F"}\}$
   - $\mathbb{P}_0(\emptyset) = 0$, $\mathbb{P}_0(\{\text{"S"}\}) = p$, $\mathbb{P}_0(\{\text{"F"}\}) = 1 - p$, $\mathbb{P}_0(\{\text{"S"}, \text{"F"}\}) = 1$
2. We have probability spaces $(S_i, \mathcal{S}_i, \mathbb{P}_i)$.
   - $S_i := \{0, 1\}$
   - Consider a Bernoulli random variable where $X_i(\text{"S"}) = 1$ and $X_i(\text{"F"}) = 0$.
   - $\mathbb{P}_i(X_i = 1) = p$ and $\mathbb{P}_i(X_i = 0) = 1 - p$.
3. We have a probability space $(S_\times, \mathcal{S}_\times, \mathbb{P}_\times)$.
   - $S_\times := \prod_i S_i$, $\mathcal{S}_\times := \prod_i \mathcal{S}_i$, and $\mathbb{P}_\times := \prod_i \mathbb{P}_i$.
4. We have a probability space $(S, \mathcal{S}, \mathbb{P})$.
   - $S := \{0, 1, \ldots, n\}$
   - Consider a new random variable $X : S_1 \times \cdots \times S_n \to S$, where $X := \sum_{i=1}^{n} X_i$ and $X_1, \ldots, X_n$ are independent and identically distributed.

Use Case: Binomial Distribution II

- $\mathbb{P}(X = m)$? Let $\mathcal{A}_m \subseteq S_1 \times \cdots \times S_n$ be a bit string with m ones.

$$
\begin{aligned}
\mathbb{P}(X = m) &= \mathbb{P}_\times\left(\bigcup_{A \in \mathcal{A}_m} \{A\}\right) \\
&= \sum_{A \in \mathcal{A}_m} \mathbb{P}_\times(\{A\}) \\
&= \sum_{A \in \mathcal{A}_m} \prod_{i=1}^{n} \mathbb{P}_i(\{A_i\}) \\
&= \sum_{A \in \mathcal{A}_m} \prod_{i=1}^{n} \mathbb{P}_0(\{\text{"S"}\} \text{ if } A_i = 1 \text{ else } \{\text{"F"}\}) \\
&= \sum_{A \in \mathcal{A}_m} p^m (1-p)^{n-m} = \binom{n}{m} p^m (1-p)^{n-m}.
\end{aligned}
$$