Probability for Statistics

Kexing Ying

May 15, 2020

Contents

1 Introduction

1.1 Probability Measures

Last year we saw briefly constructions and definitions relevant to working with probabilities such as σ-algebras, random variables and more. We will revisit them here with a more general (and more technical) approach.

Definition 1.1 (σ-algebra)**.** Let X be a set. A σ-algebra on X, $\mathcal{A}$ is a collection of subsets of X such that

- $\varnothing \in \mathcal{A}$
- for all $A \in \mathcal{A}$, $A^C \in \mathcal{A}$
- for all $(A_n)_{n=1}^{\infty} \subseteq \mathcal{A}$, $\bigcup_n A_n \in \mathcal{A}$.

Proposition 0.1. *Let X be a set and I a non-empty collection of σ-algebras on X. Then $\bigcap I$ is also a σ-algebra on X.*

This proposition is easy to check and thus, it makes sense to consider the σ-algebra generated by some set.

Definition 1.2 (Generator of σ-algebra)**.** Let X be a set and $S \subseteq \mathcal{P}(X)$ a collection of subsets of X. Then the σ-algebra generated by S is

$$\sigma(S) := \bigcap\{\mathcal{A} \supseteq S \mid \mathcal{A} \text{ is a } \sigma\text{-algebra on } X\}$$

By the fact that the power set of X is a σ-algebra containing S, we see that $\{\mathcal{A} \supseteq S \mid \mathcal{A} \text{ is a } \sigma\text{-algebra on } X\}$ is non-empty and so for all $S \subseteq \mathcal{P}(X)$, $\sigma(S)$ a (and the smallest) σ-algebra on X.

With this, we can construct a commonly seen σ-algebra, the Borel σ-algebra. Given some topological space X, the Borel σ-algebra on X is the σ-algebra generated by $\mathcal{T}_X$, i.e. $\mathcal{B}(X) = \sigma(\mathcal{T}_X)$. We will most commonly work with the Borel σ-algebra on the real numbers $\mathcal{B}(\mathbb{R})$.

We call the ordered pair $(X, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra n X a *measurable space*.

Definition 1.3 (Measure)**.** Given a measurable space $(X, \mathcal{A})$, a measure on this measurable space $\mu : \mathcal{A} \to [0, \infty]$ is a function such that

- $\mu(\varnothing) = 0$
- for all disjoint sequence $(A_n)_{n=1}^{\infty} \subseteq \mathcal{A}$, $\mu\left(\bigsqcup_n A_n\right) = \sum_n \mu(A_n)$

With measures defined, we can add an additional restriction to create a *probability space*.

Definition 1.4 (Probability Measure)**.** Let μ be a measure on the measurable space $(X, \mathcal{A})$, then μ is a probability measure if and only if $\mu(X) = 1$. We then call the order triplet $(X, \mathcal{A}, \mu)$ a probability space.

To distinguish probability space from normal measure spaces, we will often write $(\Omega, \mathcal{F}, \mathbb{P})$ to denote a probability space. We will call Ω the *sample space*, $\mathcal{F}$ the *events* and for all $A \in \mathcal{F}$, $\mathbb{P}(A)$ the *probability* of the event A.

1.2 Random variables

Now that we have the basic notion of a probability space, we would like to play around with it using *random variables*. In the most general sense, random variables are simply functions from the probability space to another measurable space, most commonly the real numbers equipped with $\mathcal{B}(\mathbb{R})$.

Definition 1.5 (Measurable Functions)**.** Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces and $f : X \to Y$ a mapping between the two. We call f measurable if and only if for all $A \in \mathcal{B}$, $f^{-1}(A) \in \mathcal{A}$.

Definition 1.6 (Random Variables)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $E, \mathcal{A}$ be a measurable space. Then an E-valued random variable is a measurable function $X : \Omega \to E$.

In general, we will only be working with real valued random variables, so the image measurable space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Often, when we have a random variable $X : \Omega \to \mathbb{R}$, we might ask questions such as "what is the probability that $X \in A$" for some $A \subseteq \mathrm{Im} X$. We now see that this question is asking for exactly $\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$ (this makes sense as X is measurable).

Another property that is useful for random variables is the notion of independence.

Definition 1.7 (Independence of Events)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(A_n) \subseteq \mathcal{F}$ a sequence of events. Then (A_n) is said to be independent if and only if for all *finite* index set I,

$$\mathbb{P}\left(\bigcap_{n \in I} A_n\right) = \prod_{n \in I} \mathbb{P}(A_n).$$

Definition 1.8 (Independence of σ-algebras)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{A}_n)$ be a sequence of sub-σ-algebras of $\mathcal{F}$. Then $(\mathcal{A}_n)$ is said to be independent if and only if for all $(A_n) \subseteq \mathcal{F}$ a sequence of events such that $A_i \in \mathcal{A}_i$, (A_n) is independent.

Equipped with these two notions of independence, it makes sense to create a notion of some σ-algebra induced by arbitrary measurable functions and with that the notion of independence of random variables is also induced.

Definition 1.9 (σ-algebra Generated by Functions)**.** Let E be a set and $\{f_i : E \to \mathbb{R} \mid i \in I\}$ be an indexed family of real-valued functions. Then the σ-algebra on E generated by these functions is

$$\sigma(\{f_i \mid i \in I\}) := \sigma(\{f^{-1}(A) \mid A \in \mathcal{B}(\mathbb{R})\}).$$

Note that with this definition, we created the smallest σ-algebra on E such that all f_i are measurable.

Definition 1.10 (Independence of Random Variables)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and (X_n) be a sequence of real-valued random variables. Then (X_n) is said to be independent if and only if the family of σ-algebras $\sigma(X_n)$ is independent.

We will check that this definition of independence of random variables behave as intended, that is $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.

Theorem 1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X, Y be real-valued random variables. THen X, Y are independent if and only if for all $A, B \in \mathcal{B}(\mathbb{R})$,*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Proof. Recall that $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}((X \in A) \cap (Y \in B)) = \mathbb{P}(X^{-1}(A) \cap Y^{-1}(B))$. Now, if $\sigma(X)$ and $\sigma(Y)$ are independent, as $X^{-1}(A) \in \sigma(X)$ and $Y^{-1}(B) \in \sigma(Y)$, by definition, we have $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$.

Similarly, if the equality in question is true for all $A, B \in \mathcal{B}(\mathbb{R})$, then the σ-algebras are independent by definition, and thus, so are the random variables. $\square$