Ma140a Probability
Lecture 5: Random variables

Juan Pablo Vigneaux

Caltech

Department of Mathematics
California Institute of Technology
Pasadena, CA, USA
vigneaux@caltech.edu

August 3, 2024

Table of Contents

1. Measurability
2. Random variables

Table of Contents

1. Measurability
2. Random variables

Inverse image

Let $f : X \to Y$ be a function between set. We define the **inverse image** of a set $A \subset Y$ as

$$f^{-1}(A) := \{\, x \in X \;:\; f(x) \in A \,\}\,.$$

Remark that this gives a function $f^{-1} : 2^Y \to 2^X$.

Example

Let $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$. Define $X : \Omega \to \mathbb{Z}$ as $X(\omega) = (\#$ of points), and $Y = X \mod 2$. Then $Y^{-1}(0) = \{$⚁, ⚃, ⚅$\}$.

Measurability

Let $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ be measurable spaces. A function $f : X \to Y$ is called **measurable** (or $\mathcal{F}/\mathcal{G}$-measurable) if for all $G \in \mathcal{G}$, $F^{-1}(G)$ belongs to $\mathcal{F}$.

Remark: In this case, there is a map $f^{-1} : \mathcal{G} \to \mathcal{F}$.

Example

Let $\Omega = \{H, T\}^{\mathbb{N}} = \{\omega : \mathbb{N} \to \{H; T\}\} = \{\,(\omega_0, \omega_1, ...)\ :\ \omega_i \in \{H, T\}\,\}$, and define $X_i : \Omega \to \mathbb{R}$ by

$$X_i(\omega) = \begin{cases} 1 & \text{if } \omega_i = H \\ 0 & \text{if } \omega_i = T \end{cases}.$$

What is the smallest σ-algebra on Ω that makes all $(X_i)_{i \in \mathbb{N}}$ measurable?
It must contain every set of the form $\{X_i = 1\}$... in fact, it is $\sigma(\{\{X_i = 1\},\, i \in \mathbb{N}\})$.

Measurability (continued)

Let $(X, \mathcal{F})$ be a measurable space. For a real-valued function $f : X \to \mathbb{R}$, it is understood that $\mathbb{R}$ is equipped with its Borel σ-algebra.

We denote by $m\mathcal{F}$ the set of real-valued measurable functions on $(X, \mathcal{F})$. In turn, $b\mathcal{F}$ denotes the subset of bounded real-valued measurable functions.

—

A function $f : X \to \mathbb{R}$ on a topological space $(X; \tau)$ is called Borel if it is $\sigma(\tau)/\mathcal{B}(\mathbb{R})$-measurable.

Functoriality

Theorem

If $f(X,\mathcal{F}) \to (Y,\mathcal{G})$ *and* $g : (Y,\mathcal{G}) \to (Z;\mathcal{H})$ *are measurable maps, then* $g \circ f : (X;\mathcal{F}) \to (Z,\mathcal{H})$ *is measurable too.*

Functoriality

Theorem

If $f(X,\mathcal{F}) \to (Y,\mathcal{G})$ *and* $g : (Y,\mathcal{G}) \to (Z;\mathcal{H})$ *are measurable maps, then* $g \circ f : (X;\mathcal{F}) \to (Z,\mathcal{H})$ *is measurable too.*

Proof.

We can compose $g^{-1} : \mathcal{H} \to \mathcal{G}$ and $f^{-1} : \mathcal{G} \to \mathcal{F}$.
In other words, for any $H \in \mathcal{H}$, $g^{-1}(H) \in \mathcal{G}$, so $f^{-1}(g^{-1}(H)) \in \mathcal{F}$, and $f^{-1}(g^{-1}(H)) = (g \circ f)^{-1}(H)$. □

There's a category whose objects are measurable spaces and whose arrows are measurable maps.

Theorem (Williams, Prop. 3.2)

Let $h : X \to Y$ be a function.

1. h^{-1} *preserves all set operations: for every index set A (of arbitrary cardinality), every collection $(E_\alpha)_\alpha \in A$, and any $E \subset Y$,*
   - $h^{-1}(\bigcup_{a \in A} E_a) = \bigcup_{a \in A} h^{-1}(E_a)$
   - $h^{-1}(E^c) = (h^{-1}(E))^c$
   - $h^{-1}(\bigcap_{a \in A} E_a) = \bigcap_{a \in A} h^1(E_a)$
2. *Suppose $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ are measurable spaces. Consider $\mathcal{C} \subset \mathcal{G}$ such that $\sigma(\mathcal{C}) = \mathcal{G}$. If for every $C \in \mathcal{C}$, $h^{-1}(C) \in \mathcal{F}$, then h is measurable.*

Proof.

(1): Exercise.

For part (2), consider the collection $\mathcal{E}$ of sets E such that $h^{-1}(E) \in \mathcal{F}$. By (1) it is a σ-algebra and by (2) it contains $\mathcal{C}$, hence it also contains $\mathcal{G}$. □

Two applications

Example

Let $Y = \mathbb{R}$ and $\mathcal{G} = \mathcal{B}(\mathbb{R})$. Then $\mathcal{G} = \sigma(\{\, (-\infty, c] \ : \ c \in \mathbb{R}\,\}$ so $h : X \to \mathbb{R}$ is measurable if $h^{-1}((-\infty, c]) \in \mathcal{F}$ for all $c \in \mathbb{R}$.

Corollary

If X, Y are topological spaces and $h : X \to Y$ is continuous, then h is also measurable.

Real-valued measurable functions form an algebra over $\mathbb{R}$

Theorem (Williams, 3.2)

For any $h_1, h_2 \in m\mathcal{F}$ *and* $\lambda \in \mathbb{R}$, *one has*

$$h_1 + h_2 \in m\mathcal{F}, \quad h_1 \cdot h_a \in m\mathcal{F}, \quad \textit{and} \quad \lambda h_1 \in m\mathcal{F}$$

Extended real line

It is convenient to introduce the extended real line $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.

This is a totally ordered set, that we equip with the order topology, which is generated by the rays $(a, +\infty]$ and $[-\infty, a)$. As a topological space, it is compact and homeomorphic to $[0, 1]$.

The rays $(a, +\infty]$ are neighborhoods of $+\infty$, and similarly the rays $[-\infty, a)$ are neighborhoods of $-\infty$.

$\mathcal{B}(\bar{\mathbb{R}})$ is the Borel σ-algebra generated by this topology.

$\pi(\bar{\mathbb{R}}) = \{\, [-\infty, c] \;:\; c \in \bar{\mathbb{R}} \,\}$ is a π-system that generates $\mathcal{B}(\bar{\mathbb{R}})$.

Measurability of limits

Theorem (Williams, 3.5)

Let $(h_n)_{n\in\mathbb{N}}$ *be q sequence of elements of* $m\mathcal{F}$*. Then* $\inf h_n$, $\sup h_n$, $\liminf h_n$ *and* $\limsup h_n$ *are measurable functions (valued in* $\bar{\mathbb{R}}$*. Further,*

$$\{\, \omega \in \Omega \; : \; \lim h_n(\omega) \text{ exists in } \mathbb{R} \,\} \in \mathcal{F}.$$

Proof.

For $\inf$: $\{\inf h_n \geq C\} = \cup_{n\in\mathbb{N}}\{h_n \geq c\}$. □

Example

Let $\Omega = \{H,T\}^{\mathbb{N}} = \{\omega : \mathbb{N} \to \{H;T\}\} = \{\,(\omega_0, \omega_1, ...)\ :\ \omega_i \in \{H,T\}\,\}$, and define $X_i : \Omega \to \mathbb{R}$ by

$$X_i(\omega) = \begin{cases} 1 & \text{if } \omega_i = H \\ 0 & \text{if } \omega_i = T \end{cases}.$$

$$\mathcal{F} = \sigma(\{\{X_i = 1\},\, i \in \mathbb{N}\}).$$

Then $S_n = X_1 + \cdots + X_n$ measurable for every $n \in \mathbb{N}^*$.
We conclude that

$$\Lambda = \{\,\omega \in \Omega\ :\ S_n(\omega)/n \to p\,\} = \{\liminf S_n/n = p\} \cap \{\limsup S_n/n = p\}$$

belongs to $\mathcal{F}$ i.e. it is measurable.

Table of Contents

1 Measurability

2 Random variables

Random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(E_X, \mathcal{F}_X)$ be a measurable set. In probabilistic terminology, a measurable function $X : (\Omega, \mathcal{F}) \to (E_X, \mathcal{F}_X)$ is an E_X**-valued random variable.**

When $E_X = \mathbb{R}$, we simply say *random variable*.

Remark

For some authors (e.g. Omer Tamuz) a random variable is an equivalence class of measurable functions, under $f \sim g$ iff $P(f \neq g) = 0$. The problem is that, under this definition, it does not make sense to talk about $f \geq 0$ or f continuous, etc. One has to go back and forth between equivalence classes and representatives.

As suggested by the examples above, Ω does not have to be a numeric space. The random variables turn the possible outcomes $\omega \in \Omega$ into numbers, that can be combined using arithmetic operations, etc.

The algebra generated by X, denoted $\sigma(X)$, is the σ-algebra generated by $X^{-1}(\mathcal{F}_X) := \{\, X^{-1}(F) \,:\, F \in \mathcal{F}_X \,\}$. This is the smallest σ-algebra that makes X measurable. More generally:

Definition

Given a collection $(Y_\alpha : \Omega \to (E_\alpha, \mathcal{F}_\alpha),\ \alpha \in A)$ of maps on a set Ω into measurable spaces $(E_\alpha, \mathcal{F}_\alpha)$,

$$\sigma(Y_\alpha : \alpha \in A) := \sigma(\bigcup_{a \in A} Y_\alpha^{-1}(\mathcal{F}_\alpha))$$

is the smallest σ-algebra that makes all Y_α measurable simultaneously.

$\sigma(Y_\alpha : \alpha \in A)$ contains all the events that can be defined in terms of Y_αs, such that $\{a \leq Y_\alpha \leq b, c \leq Y_\beta \leq d\}$, etc.

Law and distribution function

$(\Omega, \mathcal{F}, P)$ probability triple, $X : \Omega \to (E_X, \mathcal{F}_X)$.

$$\mathcal{F}_X \xrightarrow{\quad X^{-1} \quad} \mathcal{F} \xrightarrow{\qquad P \qquad} [0,1]$$

$$A \longmapsto \{X \in A\} \longmapsto P(X \in A)$$

Remember that $\{X \in A\} = \{\omega \in \Omega \,:\, X(\omega) \in A\}$.

The law of X is the measure $L_X = P \circ X^{-1}$ defined on $(E_X, \mathcal{F}_X)$. It is a probability measure.

When $E_X = \mathbb{R}$, we saw that L_X is uniquely determined by $F_X : \mathbb{R} \to [0,1]$ such that $F_X(x) = L_X((-\infty, x]) = P(X \leq x)$.

F_X is called the **cumulative distribution function** (cdf) of X.

Theorem (Williams 3.10)

Suppose F is the distribution function of some real-valued random variable. Then

1. $F : \mathbb{R} \to [0, 1]$ *is nondecreasing.*
2. $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to +\infty} F(x) = 1$.
3. F *is right continuous.*

Proof.

The first follows form the monotonicity of P under inclusions, the second from the monotone-convergence properties of measures, and the third because

$$P(X \leq x + 1/n) \downarrow P(X \leq x).$$

□

Given a function F that satisfies this properties, one can construct an X whose cdf is F (Skorokhod representation, Williams 3.12).

Discrete variables

Suppose F_X is piece-wise constant, with discontinuities at $\Lambda = \{x_1, x_2, ...\}$ (exercise: there is at most a countable number of such discontinuities).

For $x \in \Lambda$, set

$$p(x) = F(x) - \lim_{x \to x^-} F(x).$$

Then $\sum_{x \in \lambda} p(x) = 1$. This p is called a **probability mass function.**

Example

$\Lambda = \{0, 1, 2, ..., n\}$ and $p(k) = \binom{n}{k} p^k (1-p)^k$ defines a **binomial** distribution of parameter $p \in [0, 1]$.

Example

$\Lambda = \mathbb{N} = \{0, 1, 2, ...\}$ and $p(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ defines a **Poisson** distribution of parameter $\lambda > 0$.

Continuous variables

Suppose that $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$ for some integrable function f.

By monotonicity of F, f must be positive.

And since $\lim_{x\to+\infty} F(x) = 1$, $\int_{-\infty}^{+\infty} f(x) = 1$.

f is called a probability density function.

Example

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$ defines a **normal** distribution of mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$.

Example

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

defines an **exponential** distribution of parameter $\lambda > 0$.