NMAI059 – Probability and Statistics 1

Mykhaylo Tyomkyn

Lecture 6 - Continuous measures and random variables.

For countable sample spaces Ω we use $\mathcal{F} = 2^\Omega$. On the other hand, for uncountable sample spaces this would have some drawbacks. As the existence of Vitali sets demonstrates, we would no longer be able to define a translation invariant measure, i.e. a probability function $\mathbb{P}$ such that for every fixed $t \in \mathbb{R}$ and $A \subset \mathbb{R}$ we have $\mathbb{P}(A+t) = \mathbb{P}(A)$, where $A + t := \{x + t : x \in A\}$. To remedy this we restrict the event space. This turns out not to be a major problem, as $2^\mathbb{R}$ is 'too big' anyway: most subsets of $\mathbb{R}$ are too chaotic and uninteresting for any practical or even theoretical purpose. So, what should serve as $\mathcal{F}$ when $\Omega = \mathbb{R}$? It is the so-called *Lebesgue σ-algebra* $\mathcal{L} = \mathcal{L}(\mathbb{R})$. Defining it properly would require delving into the area of mathematics known as *measure theory*, for which we do not have the time. We will therefore introduce $\mathcal{L}$ and the underlying *Lebesgue measure* λ matter of factly.

Fact 1 (Lebesgue σ-algebra) [1] *There exists a set family $\mathcal{L} \subseteq 2^\mathbb{R}$ with the following properties, members of $\mathcal{L}$ are called* (Lebesgue) measurable sets.

1. *$\emptyset \in \mathcal{L}$ and $\mathbb{R} \in \mathcal{L}$.*
2. *$A \in \mathcal{L} \Rightarrow \mathbb{R} \setminus A \in \mathcal{L}$.*
3. *$A_1, A_2, \ldots \in \mathcal{L} \Rightarrow \bigcup_{i=1}^\infty A_i \in \mathcal{L}$. That is, countable unions of measurable sets are measurable. Note that by 2. and De Morgan's law, countable intersections are also measurable.*
4. *$\mathcal{L}$ contains all closed and all open intervals. More generally, $\mathcal{L}$ contains all open and closed sets on $\mathbb{R}$ as the metric space (with the standard metric).*
5. *$\mathcal{L}$ is translation and dilation invariant. That is, for every $A \in \mathcal{L}$ and $x, y \in \mathbb{R}$ we have $xA+y \in \mathcal{L}$, where $xA + y = \{xa + y \colon a \in A\}$.*

In summary, $\mathcal{L}$ contains essentially every subset of $\mathbb{R}$ imaginable.

Exercise 1 *Show that the Cantor set belongs to $\mathcal{L}$.*

The above list of properties of $\mathcal{L}$ would by itself be vacuous, as $2^\mathbb{R}$ of course has all of them. What makes it crucial is the existence of a translation invariant measure.

Fact 2 (Lebesgue measure) *On $\mathcal{L} = \mathcal{L}(\mathbb{R})$ there exists a function $\lambda \colon \mathcal{L} \to [0, \infty]$ (where we allow $+\infty$ as a value) known as the* Lebesgue measure *on $\mathbb{R}$. It has the following properties.*

1. *$\lambda(\emptyset) = 0$ and $\lambda(\mathbb{R}) = +\infty$.*

[1] In general, a *σ-algebra* on a set Ω is a family $\mathcal{F} \subseteq 2^\Omega$ satisfying conditions 1.-3. below, with Ω in place of $\mathbb{R}$.

2. $\lambda([a,b]) = b - a$ *for all* $a \leq b$.
3. λ *is* σ*-additive. In particular,* $\lambda(I) = +\infty$ *for every unbounded interval* I.
4. *For every* $A \in \mathcal{L}$ *and* $x, y \in \mathbb{R}$ *we have* $\lambda(xA + y) = |x|\lambda(A)$.
5. *For every* $A \in \mathcal{L}$ *with* $\lambda(A) = 0$ *(we say* A *is a* null set*) we have* $2^A \subseteq \mathcal{L}$ *and* $\lambda(B) = 0$ *for every* $B \subseteq A$.

Note that λ is not a probability measure, as $\lambda(\mathbb{R}) = +\infty \neq 1$. However, we can turn it into a probability measure if we restrict and rescale it to a bounded interval.

Definition 1 (Continuous uniform measure) *For real numbers* $a < b$ *the* uniform probability measure *on*[2] $\Omega = [a,b]$*, in notation* $Unif([a,b])$*, is the function* $\mathbb{P} : \mathcal{L} \cap 2^{\Omega} \to [0,1]$,

$$\mathbb{P}(A) = \frac{\lambda(A)}{\lambda(\Omega)} = \frac{\lambda(A)}{b-a}.$$

Note that, assuming Fact 2, this is indeed a probability measure satisfying Kolmogoroff's axioms.

Remark 1 *The notions of* $\mathcal{L}$*,* λ *and the uniform probability measure can be extended to* $\mathbb{R}^n$*, where the Lebesgue measure of a bounded set corresponds to its 'area' or 'volume'.*

Definition 2 (Lebesgue measurable function) *A function* $f : \mathbb{R} \to \mathbb{R}$ *is a* (Lebesgue) measurable function *if for every* $A \in \mathcal{L}$ *we have* $f^{-1}(A) \in \mathcal{L}$*, where* $f^{-1}(A) = \{x \in \mathbb{R} : f(x) \in A\}$ *(the* pre-image*).*

In practice, every continuous or otherwise 'normally defined' function is measurable. In fact, defining a non-measurable functions equates in hardness to defining a non-measurable set.

Fact 3 (Lebesgue integral) *The notion of the Riemann integral* $\int_a^b f(x)dx$*, defined initially for continuous functions, can be extended to all Lebesgue measurable functions* $f : \mathbb{R} \to [0, \infty)$*, defining the* Lebesgue integral $\int f d\lambda \in \mathbb{R} \cup \{+\infty\}$ *such that*[3]

1. *When* f *is Riemann integrable, its Riemann and Lebesgue integrals have the same value*
2. *For any sequence* $f_1, f_2, \ldots$ *of non-negative measurable functions we have*

$$\int (\sum_{n \geq 1} f_n) d\lambda = \sum_{n \geq 1} (\int f_n d\lambda)$$

Note that the Riemann integral, apart from being defined for a narrower set of functions, does not have property 2, as $\sum_{n \geq 1} f_n$ may no longer be Riemann integrable. Hence, the Lebesgue integral is superior to the Riemann integral in every possible regard.

Furthermore, the Lebesgue integral makes it possible to use measurable functions as so-called *probability density functions* in order to define further continuous probability measures.

[2] The uniform measure on an open or half-open bounded interval is defined in exactly the same way.

[3] It can be extended further to functions $f : \mathbb{R} \to \mathbb{R}$, unless the positive and negative parts both have unbounded integrals

Definition 3 (Continuous measures) *Let D be a measurable subset of $\mathbb{R}$ (typically, $D = \mathbb{R}$ or D is an interval) and $f\colon D \to [0,\infty)$ be a measurable function with $\int f d\lambda = 1$. Define the probability measure $\mathbb{P}_f$ on $\mathcal{L} \cap 2^D$ by*

$$\mathbb{P}_f(A) = \int_A f d\lambda = \int f \cdot \mathbb{1}_A d\lambda,$$

where $\mathbb{1}_A$ is the indicator function of A, i.e. $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise. The function f is the probability density function *of $\mathbb{P}_f$.*

Example 1 *For $a < b$, using $f\colon [a,b] \to \mathbb{R}$, $f(x) = 1/(b-a)$, we recover the uniform measure on $[a,b]$ from the previous example.*

Example 2 (Exponential distribution) *Let $\lambda > 0$ and let*[4]

$$f(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{[0,\infty)}(x).$$

Since $\int_{\mathbb{R}} f(x)dx = 1$, this defines a continuous probability distribution, the Exponential distribution *with parameter λ, denoted $Exp(\lambda)$. This is the continuous analogue of the geometric distribution.*

Example 3 (Cauchy distribution) *Based on the fact that $\int_{-\infty}^{\infty} \frac{1}{x^2+1}dx = \pi$, we can define a probability measure by the density function $f(x) = \frac{1}{\pi(1+x^2)}$.*

The next example is the single most important distribution in probability theory and statistics.

Example 4 (Normal/Gaussian distribution) *Based on the fact that $\int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}$, the function*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

defines a probability measure on $(\mathbb{R}, \mathcal{L})$, the Normal *or* Gaussian distribution *$\mathcal{N}(0,1)$. More generally, for parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, the distribution $\mathcal{N}(\mu,\sigma^2)$ is defined by the density function*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Example 5 (Gamma distribution) *The* Gamma function *$\Gamma : (0,\infty) \to \mathbb{R}$ is defined as $\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x}dx$. In particular, $\Gamma(1) = 1$ and, integrating by parts: $\Gamma(t+1) = t\Gamma(t)$. So, for an integer t we have $\Gamma(t) = (t-1)!$ (the Gamma function is the 'continuous factorial'). This gives rise to the distribution $Gamma(t,\lambda)$ with $t, \lambda > 0$, via the density function*

$$f(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} \cdot \mathbb{1}_{[0,\infty)}(x).$$

Note that $\Gamma(1,\lambda)$ is the exponential distribution $Exp(\lambda)$. This is not a coincidence, we will return to it when discussing the (continuous) convolution formula.

Definition 4 (General random variable) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A (real-valued)* random variable *is a function $X\colon \Omega \to \mathbb{R}$ such that $X^{-1}(L) \in \mathcal{F}$ for all $L \in \mathcal{L}$.*

[4] λ here denotes a real number, not the Lebesgue measure

Exercise 2 *Verify that a discrete random variable, as defined previously, is a random variable.*

Definition 5 (Distribution of a random variable) *The* distribution *or* pushforward measure *of* $X\colon(\Omega,\mathcal{F},\mathbb{P})\to\mathbb{R}$ *is the probability measure* $\mathbb{P}_X$ *on* $(\mathbb{R},\mathcal{L})$ *defined by*

$$\mathbb{P}_X(L)=\mathbb{P}(X^{-1}(L))$$

Definition 6 (Continuous random variable) *A random variable X is* continuous *if $\mathbb{P}_X$ is continuous. In that case the probability density function of $\mathbb{P}_X$ is referred to as* probability density function (pdf) *of X, and denoted f_X.*

As in the discrete case, we will mainly care about the distribution and the probability density function of a continuous random variable, rather than its origin.

Definition 7 (Cumulative distribution function for general random variables) *Let $X:\Omega\to\mathbb{R}$ be a random variable. Then*

$$F_X:\mathbb{R}\to[0,1],\quad F_X(x):=\mathbb{P}(X\leq x)$$

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

Remark 2 *A random variable X can be*

- *discrete (have a probability mass function),*
- *continuous (have a probability density function),*
- *neither. An example would be $X=Y+Z$, where Y is discrete and Z is continuous. However, there even exist 'singular' random variables that do not have any discrete or continuous 'components'.*[5]

However, F_X exists always.

The basic properties of F_X are the same as in the discrete case: we have $\lim_{x\to-\infty}F_X(x)=0$, $\lim_{x\to+\infty}F_X(x)=1$, F_X is non-decreasing and right-continuous[6].

Theorem 1 (Relationship between pdf and cdf) *Let X be a continuous random variable. Then*

$$F_X(x)=\int_{-\infty}^{x}f_X(t)dt,\text{ and if } f_X \text{is continuous then } f_X=\frac{dF_X}{dx}.$$

Example 6 *For $X\sim Exp(\lambda)$ and $x\geq 0$ we have*

$$F_X(x)=\mathbb{P}(X\leq x)=\int_0^x\lambda e^{-\lambda t}dt=1-e^{-\lambda x}.$$

[5]That being said, all measures and random variables we will be dealing with in this course will be either discrete or continuous.

[6]F_X is (both right and left) continuous when X is continuous