Probability theory

Kristian Wichmann

October 30, 2016

This is an overview of probability theory expressed in the language of measure theory.

1 Probability spaces

Definition 1.1. *A probability space is a measure space* $(\Omega, \mathcal{F}, P)$ *for which* $P(\Omega) = 1$*. The elements of* Ω *are called outcomes, the elements of* $\mathcal{F}$ *events and* $P(A)$ *is the probability of the event* $A \in \mathcal{F}$*.*

Note that we can trivially express the probability of the event $A \in \mathcal{F}$ happening as:

$$P(A) = \int_A dP \tag{1}$$

2 Random variables

Definition 2.1. *An* E*-valued random variable* X *on a probability space* $(\Omega, \mathcal{F}, P)$ *is a measurable function* $X : \Omega \to E$*. Here* $(E, \mathcal{E})$ *is a measurable space. Often this measurable space is* $(\mathbb{R}, \mathbb{B})$*, so that* $X : \Omega \to \mathbb{R}$*. Such a real-valued random variable is usually simply denoted a random variable for brevity.*

2.1 Distribution and expectation values

Consider a random variable X as described above. Since X is a measurable function, it induces an image measure $P(X)$ on E. This is also sometimes known as a *pushforward measure*, or the *distribution*. This measure is used in the following definition:

Definition 2.2. *The expectation value of X (if it exists) is given by:*

$$E[X] = \int_\Omega X(\omega)\ dP(X) \tag{2}$$

2.2 Moments

If the random variable X is real-valued, we may make the following definition:

Definition 2.3. *The n'th moment of a real-valued random variable X is:*

$$m_n = E[X^n] \tag{3}$$

Here, X^n is the function $X^n : \omega \mapsto (X(\omega))^n$ as usual.

2.3 Distribution with respect to other measures

Let's start this section with a reminder from measure theory:

2.3.1 Absolute continuity and the Radon-Nikodym theorem

Consider two measures μ and ν on the measurable space $(X, \mathbb{E})$. Then ν is said to be *absolutely continuous* with respect to μ is all null sets of μ are also null sets of ν. We also say ν is *dominated* by μ and write $\nu \ll \mu$.

Theorem 2.1. *(Radon-Nikodym) Let $\nu \ll \mu$. If μ is σ-finite, then there exists a function $f : X \to \mathbb{R}$ such that:*

$$\forall\ A \in \mathbb{E} :\ \nu(A) = \int_A f\ d\mu \tag{4}$$

The function f is known as the *Radon-Nikodym derivate* of ν with respect to μ and is sometimes written:

$$f = \frac{d\nu}{d\mu} \tag{5}$$

2.3.2 Application to probability theory

We now return to the case of X being a random variable on the probability space $(\Omega, \mathcal{F}, P)$.

Let μ be a σ-finite measure on Ω so that $P \ll \mu$. This is known as a *dominating measure.* We can now use theorem 2.1 to write:

$$\forall\ A \in \mathcal{F} :\ P(A) = \int_A f\ d\mu \tag{6}$$

Here, $f = \frac{dP}{d\mu}$, the Radon-Nikodym derivative of P with respect to μ. Comparing to equation (1) we might think of this as a "change of variable", and it is now natural to say:

Definition 2.4. *If μ is a σ-finite measure, dominating P, then the Radon-Nikodym derivative with respect to P is called the probability density function with respect to μ:*

$$f_\mu = \frac{dP}{d\mu} \tag{7}$$

Specifically note, that the probability density function with respect to P itself is just the constant unit function. The probability density function is sometimes abbreviated as *pdf*. In practice, the dominating measure is usually the Lebesgue measure, the counting measure, or a combination of the two.

3 Statistical models

4 Likelihood