Infotheory for Statistics and Learning

Lecture 0

- Basic concepts in probability

Measurable space $(\Omega, \mathcal{A})$, with $\mathcal{A} =$ the class of measurable sets

- $\mathcal{A}$ is a σ-algebra, i.e. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$ and $A_i \in \mathcal{A} \Rightarrow \cup_i A_i \in \mathcal{A}$

Measure space $(\Omega, \mathcal{A}, \mu)$ with $\mu(A)$ the measure of $A \in \mathcal{A}$

Class of Borel sets on $\mathbb{R}$ = smallest σ-algebra that contains the open intervals (all sets of the form (a, b), $b > a$), notation $\mathcal{B}(\mathbb{R})$

A real-valued function $f : \Omega \to \mathbb{R}$ is measurable if

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \text{ for all } B \in \mathcal{B}$$

For fixed $n < \infty$, $a_i \in \mathbb{R}$ and $A_i \in \mathcal{A}$, $i = 1, \ldots, n$, a measurable function of the form

$$s(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x)$$

(with $\chi_A(x) = 1$ if $x \in A$ and $= 0$ o.w.) is called a simple function

The integral of a simple function $s(x)$ is defined as

$$\int s(x)d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$$

Any nonnegative measurable function $f(x)$ can be written as $f(x) = \lim_{i\to\infty} s_i(x)$ where $\{s_i(x)\}$ is a sequence of nonnegative simple functions such that $s_1(x) \le s_2(x) \le \cdots \le f(x)$

The integral of a nonnegative measurable f is then obtained as

$$\int f(x)d\mu = \lim_{i\to\infty} \int s_i(x)d\mu$$

The integral of a general measurable function is defined as

$$\int f(x)d\mu = \int \max\{f(x),0\}d\mu - \int (-\min\{f(x),0\})d\mu$$

A measurable function f is integrable if $\int |f|d\mu < \infty$

Probability space $=$ a measure space $(\Omega, \mathcal{A}, P)$ such that $P(\Omega) = 1$

Random variable $=$ a measurable real-valued function $X : \Omega \to \mathbb{R}$

Distribution P_X of X, for $E \in \mathcal{B}$,

$$P_X(E) = P(\{\omega : X(\omega) \in E\})$$

- We get a new probability space $(\mathbb{R}, \mathcal{B}, P_X)$

Probability distribution $F_X(x)$ of X,

$$F_X(a) = P_X(\{X \le a\}) = P(\{\omega : X(\omega) \le a\})$$

Expectation, for measurable $g(x)$

$$E[g(X)] = \int_{\mathbb{R}} g(x)dP_X = \int_{\Omega} g(X(\omega))dP$$

X is discrete if there is a countable $K \in \mathcal{B}$ such that $P_X(K) = 1$

For a discrete X we define $p_X(k) = P(\{\omega : X(\omega) = k\})$, $k \in K$, the probability mass function (pmf) of X

X is continuous if $P_X(\{x\}) = 0$ for all $x \in \mathbb{R}$

X is absolutely continuous if there is a function $f_X(x) \geq 0$ such that

$$P_X(B) = \int_B f_X(x)dx = \int \chi_B(x) f_X(x)dx$$

for all $B \in \mathcal{B}$, the probability density function (pdf) of X

Integration/expectation:

If X is discrete on K

$$E[g(X)] = \int g(x)dP_X = \sum_{k \in K} g(k)p_X(k) = \sum_x g(x)p_X(x)$$

If X is absolutely continuous

$$E[g(X)] = \int g(x)dP_X = \int g(x)f_X(x)dx$$

Assume that P and Q are two probability measures on $(\Omega, \mathcal{A})$, and that $P(A) = 0$ for all $A \in \mathcal{A}$ where $Q(A) = 0$, notation $P \ll Q$

Then there exists a measurable function f such that

$$P(B) = \int_B f(\omega) dQ$$

for all $B \in \mathcal{A}$

The function f is called the Radon–Nikodym derivative of P w.r.t. Q, notation

$$f(\omega) = \frac{dP}{dQ}(\omega)$$

Note that f is a random variable

On $(\mathbb{R}, \mathcal{B})$:

For two discrete distributions P_X and Q_X with pmfs p and q

$$\frac{dP_X}{dQ_X}(x) = \frac{p(x)}{q(x)}$$

For two absolutely continuous distributions P_X and Q_X with pdfs f and g

$$\frac{dP_X}{dQ_X}(x) = \frac{f(x)}{g(x)}$$

Note that if λ is Lebesgue measure[1] on $\mathcal{B}$ and $P_X \ll \lambda$ then

$$P_X(B) = \int_B \frac{dP_X}{d\lambda}(x) d\lambda$$

That is,

$$f_X(x) = \frac{dP_X}{d\lambda}(x)$$

is the pdf of P_X

Thus, X is absolutely continuous $\iff P_X \ll \lambda$

---

[1]With a more general definition of R–N derivative, since Lebesgue measure is not a prob. measure