

## Probability.

---

### Main idea:

$(\Omega, \mathcal{F}, P)$ ,  $P(\Omega) = 1$  a probability space.

If picking  $n$  points,  $\{\omega_{n,k}\}_{k=1}^n$  “at random” from  $\Omega$ , so all  $\omega_{n,k} \in \Omega$ , then the following will be true

$$\lim_{n \rightarrow \infty} \frac{\#\{k \in \{1, 2, \dots, n\}; \omega_{n,k} \in E\}}{n} = P(E), \text{ for all } E \in \mathcal{F},$$

where  $\#$  is the counting measure.

### Nonsense:

$\alpha : \mathbb{N} \rightarrow \Omega$ ,  $n \in \mathbb{N}$ , onto, but not one to one. Define  $\#_n = \frac{1}{n}\#$ ,  $N = \{1, 2, \dots, n\}$ . Then  $(N, \mathcal{P}(N), \#_n)$  is a probability space. Let  $\alpha_n = \alpha|_N$ , then this is an  $\Omega$  valued random variable. Now we can define

$$P(E) = \lim_{n \rightarrow \infty} \#_n \alpha_n^{-1}(E), \text{ for all } E \in \mathcal{F}$$

### Random variables:

$(S, \mathcal{S})$  a measurable space,  $X : \Omega \rightarrow S$  is called a random variable when it is  $(\mathcal{F}, \mathcal{S})$ -measurable.

Define  $P_X : \mathcal{S} \rightarrow [0, +\infty]$  by  $P_X(E) = P(\{\omega \in \Omega; X(\omega) \in E\})$ , this is the image measure by  $X$ .

$$P_X(S) = P(\{\omega \in \Omega; X(\omega) \in S\}) = P(\Omega) = 1$$

So the image measure induced by a random variable is a probability measure on its state space.

$P_X$  is called the distribution of  $X$ .

Define  $F_X : \mathbb{R} \rightarrow [0, 1] = x \mapsto P_X((-\infty, x])$ , this is called the cumulative distribution function.

From wikipedia:

“The probability density function of a random variable is the RadonNikodym derivative of the induced measure with respect to some base measure (usually the Lebesgue measure for continuous random variables).”

ADD many details here

### Expectation:

Define the expectation value of  $X$  as  $E(X) = \int_{\Omega} X dP$ , the integral of  $X$ .

Suppose  $X$  is a simple function, then  $X(\omega) = \sum_{k=1}^n c_k \chi_{E_k}(\omega)$ ,  $c_k \in S$ , unique, and  $E_k \in \mathcal{F}$  disjoint.

$$E(X) = \sum_{k=1}^n c_k P(E_k)$$

### Discrete rv:

A discrete random variable  $X$  is one whose state space is countable. In this case there is a bijective map,  $\gamma : S \rightarrow \mathbb{N}$ , and clearly the function  $\gamma \circ X$  is  $(\mathcal{F}, \mathcal{P}(\mathbb{N}))$ -measurable. We may write  $S = \{x_k := \gamma^{-1}(k)\}_{k=1}^{\infty}$ , and may define  $E_k := X^{-1}(x_k)$ ,  $X(\omega) = \sum_{k=1}^{\infty} x_k \chi_{E_k}$ . If we temporarily adopt the notation " $p(x_k) = P(X = x_k) := P(E_k)$ ", then

$$E(X) = \sum_{k \in \mathbb{N}} x_k p(x_k)$$

In this simple case  $\Omega$  may not really be necessary, as  $(\{x_k\}, \mathcal{P}(\{x_k\}), p)$  is a probability space in its own right, and note, with  $\beta_n := X \circ \alpha_n$ ,  $\alpha_n$  as in the above nonsense,

$$p(x_k) = \lim_{n \rightarrow \infty} \#_n \beta_n^{-1}(x_k), \text{ for all } x_k$$

## Conditional Expectation

$(X, \mathcal{A}, \mu)$  a  $\sigma$ -finite measure space, and  $(X, \mathcal{F})$  a measurable space,  $\mathcal{F} \subset \mathcal{A}$ . For any  $f : X \rightarrow \mathbb{R}$ ,  $\mathcal{A}$ -measurable,  $h : X \rightarrow \mathbb{R}$  is the conditional expectation of  $f$  with respect to  $\mathcal{F}$  if it is  $\mathcal{F}$ -measurable and for all  $A \in \mathcal{F}$ ,

$$\int_A f d\mu = \int_A h d\mu,$$

we write  $E[f|\mathcal{F}]$  for the conditional expectation of  $f$  w.r.t.  $\mathcal{F}$ .

If in addition,  $f \in L^1(X, \mathcal{A}, \mu)$ , then  $\nu \in \mathcal{M}(X, \mathcal{A})$ , where  $\nu(E) := \int_E f d\mu$ .  $(X, \mathcal{F}, \mu|_{\mathcal{F}})$  is a measure space,  $\nu|_{\mathcal{F}} \in \mathcal{M}(X, \mathcal{F})$ , and  $\nu|_{\mathcal{F}} \ll \mu|_{\mathcal{F}}$ , then by the RNT, there exists a unique  $h \in L^1(X, \mathcal{F}, \mu|_{\mathcal{F}})$  such that  $\nu(E) = \int_E h d\mu|_{\mathcal{F}}$ .

$$E[f|\mathcal{F}] = \frac{d\nu}{d\mu|_{\mathcal{F}}}, \text{ where } \nu(E) = \int_E f d\mu \text{ for all } E \in \mathcal{A}, f \in L^1(X, \mathcal{A}, \mu), \mathcal{F} \subset \mathcal{A}.$$

Remarks:

1. For  $(X, \mathcal{A}, \mu)$  a measure space,  $f, g : X \rightarrow \mathbb{R}$ ,  $\mathcal{A}$ -measurable. Then  $\int_A f d\mu = \int_A g d\mu$  for all  $A \in \mathcal{A} \Rightarrow f = g$ ,  $\mu$ -a.e. Proof: let  $h = f - g$ ,  $A^+ = h^{-1}([0, \infty])$ ,  $A^- = h^{-1}((-\infty, 0])$ . Then  $h \geq 0$  on  $A^+$  so  $\int_{A^+} h d\mu = 0 \Rightarrow h = 0$   $\mu$ -a.e. on  $A^+$  by the vanishing principle. Similarly,  $h^- = 0$   $\mu$ -a.e. on  $A^-$ .
2. In (1), what is really needed for the result is that the integrals of  $f$  and  $g$  agree on  $(f - g)^{-1}(\mathbb{R}^+) \in \mathcal{A}$  and  $(f - g)^{-1}(\mathbb{R}^-) \in \mathcal{A}$ .
3. Let  $h$  be the conditional expectation of  $f$  w.r.t.  $\mathcal{F}$ , with  $F \subset \mathcal{A}$ ,  $\Delta := f - h$ .  $h$  is  $\mathcal{F}$ -measurable, so is  $\mathcal{A}$ -measurable, so  $\Delta$  is too, and so  $\Delta^{-1}(\mathbb{R}^{+,0}) \in \mathcal{A}$ , and  $\Delta^{-1}(\mathbb{R}^-) \in \mathcal{A}$ . By (2), if in addition,  $\Delta^{-1}(\mathbb{R}^{+,0}) \in \mathcal{F}$ , and  $\Delta^{-1}(\mathbb{R}^-) \in \mathcal{F}$ , then  $f = h$   $\mu$ -a.e., so in some sense, the larger  $\mathcal{F}$  is, the closer an approximation  $h$  is of  $f$ .