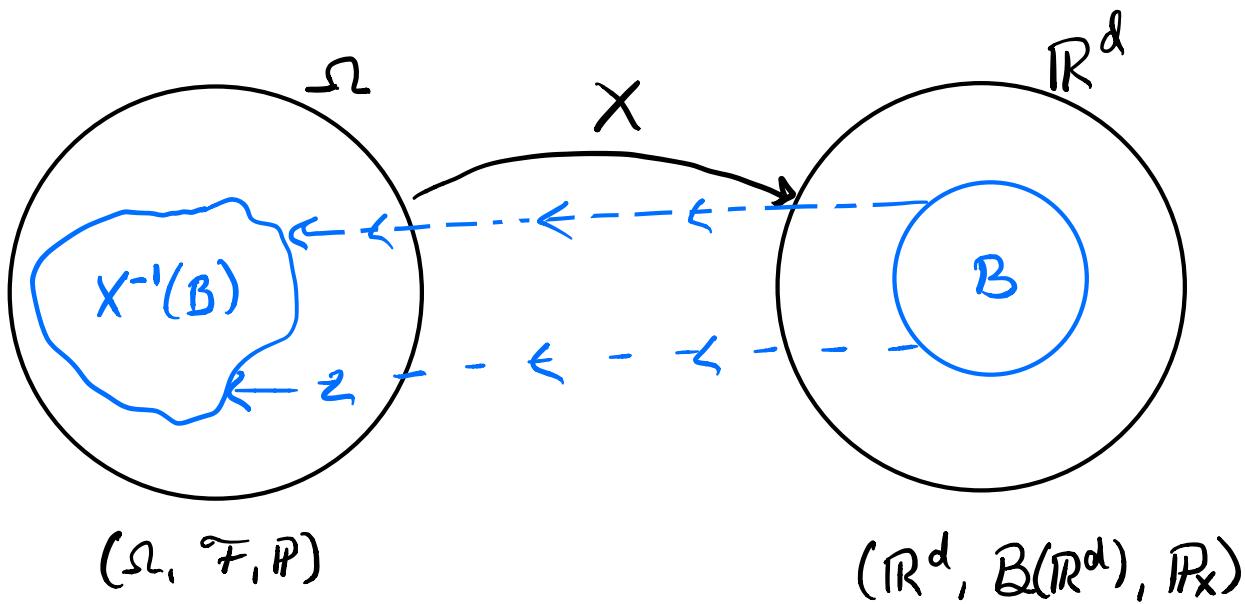


Recap

Random Variables: X is a r.v. on (Ω, \mathcal{F}, P)
 if $X: \Omega \rightarrow \mathbb{R}^d$ with $X^{-1}(B) \in \mathcal{F}$ for $B \in \mathcal{B}(\mathbb{R}^d)$



where

$$P_X(B) \triangleq P(X^{-1}(B)) \quad B \in \mathcal{B}(\mathbb{R}^d)$$

$\underbrace{\phantom{X^{-1}(B)}}_{\in \mathcal{F}}$

CDF

Define

$$F_X: \mathbb{R}^d \rightarrow [0, 1]$$

by

$$F_X(a_1, \dots, a_d) \triangleq P_X((-\infty, a_1] \times \dots \times (-\infty, a_d])$$

Properties of CDFs

Let $F_X(x)$ be the cumulative distribution function of a real-valued random variable X (*i.e.* $X: \Omega \rightarrow \mathbb{R}$, $d=1$). Then

$$(i) \lim_{t \rightarrow \infty} F_X(t) = 1$$

$$(ii) \lim_{t \rightarrow -\infty} F_X(t) = 0$$

(iii) F_X is monotonically non-decreasing and right continuous.

$$(iv) P_X((a, b]) = F_X(b) - F_X(a)$$

Discrete vs. Continuous Random Variables

Definition (support): The support IP_X denoted by $\text{Supp}(\text{IP}_X)$ is the smallest Borel set $B \in \mathcal{B}(\mathbb{R}^d)$ s.t. $\text{IP}_X(B) = 1$.

① Discrete Random Variables: X is discrete when $\text{Supp}(\text{IP}_X)$ is countable.

Then we can define the pmf

$$P_X: \text{Supp}(\text{IP}_X) \rightarrow [0, 1]$$

associated with X .

$$p_X(x) \triangleq \text{IP}_X(\{x\}) = \lim_{\varepsilon \rightarrow 0} F_X(x) - F_X(x - \varepsilon), \quad x \in \text{Supp}(\text{IP}_X)$$

② Continuous Random Variables: When $\text{Supp}(P_X)$ is uncountable AND F_X is differentiable on the interior of the support, then X is a continuous random variable.

The probability density function (pdf) of X is

$$f_X(t) = \frac{dF_X(t)}{dt}$$

$$\Rightarrow P_{f_X} = P_X$$

Expectation, Variance, Etc

Expectation: Let X be a rv on (Ω, \mathcal{F}, P) .
The expectation of X is

$$\mathbb{E}[x] \triangleq \int_{\Omega} x dP_X(x)$$

where

① if X is discrete

$$\mathbb{E}[x] = \sum_{x \in \text{Supp}(P_X)} x \cdot p_X(x)$$

② if X is continuous

$$\mathbb{E}[x] = \int_{\text{supp}(P_x)} x \cdot f_x(x) dx$$

Proposition - Expectation of a Function

Let

$$X: \Omega \rightarrow \mathbb{R}^d$$

be a random variable on (Ω, \mathcal{F}, P) and

$$g: \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$$

be a measurable function.

Then

$$\mathbb{E}[g(X)] \triangleq \int_{\mathbb{R}^d} g(x) dP_x(x)$$

or $\text{supp}(P_x)$

Variance: Let X be a r.v. on (Ω, \mathcal{F}, P) . The variance of X is

$$\text{Var}(X) \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2]$$

just think of this as $g(X)$!

Law of Large Numbers

Let X_1, X_2, \dots be a sequence of \mathbb{R} -valued iid r.v.'s on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

① Weak LLN

If $\mathbb{E}[|X_i|] < \infty \forall i$, then

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^{\infty} X_i - \mathbb{E}[X_i]\right| > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$$

② Weak LLN for Functions

Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a measurable function such that $\mathbb{E}[f(X_i)] < \infty \forall i$, then

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}[f(X_i)]\right| > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$$

③ Uniform Weak LLN

Let $f_1, \dots, f_k: \mathbb{R} \rightarrow \mathbb{R}$ be measurable functions with $\mathbb{E}[f_k(X_i)] < \infty \forall k \in \{1, \dots, k\}$.

Then

$$\mathbb{P}\left(\bigcup_{l=1}^k \left\{ \left| \frac{1}{n} \sum_{i=1}^n f_l(X_i) - \mathbb{E}[f_l(X_i)] \right| \right\} > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0 \quad \forall \varepsilon > 0$$

Conditional Probability and Transition Kernels

Transition Kernel: Let (X, \mathcal{F}) and (Y, \mathcal{G}) be two measurable spaces. A transition kernel

$$K(\cdot | \cdot) : \mathcal{G} \times X \rightarrow \mathbb{R}$$

such that

$$\textcircled{1} \quad K(\cdot | x) \in \mathcal{P}(Y), \quad \forall x \in X$$

set of ALL probability measures over target space Y

$$\textcircled{2} \quad K(B | \cdot) : X \rightarrow \mathbb{R} \text{ is } \mathcal{F}\text{-measurable}$$

(i.e. a random variable wrt (X, \mathcal{F}) , $\forall B \in \mathcal{G}$)

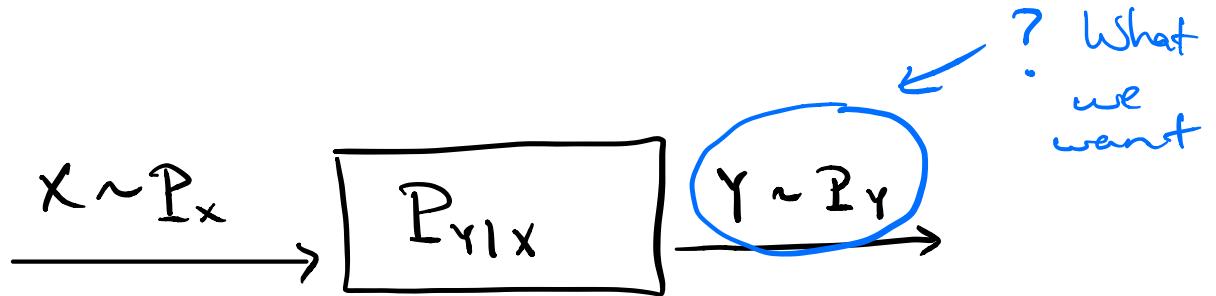
Aside: Notation

We will typically use $P_{Y|X}$ for a transition kernel from (X, \mathcal{F}) to (Y, \mathcal{G}) , to stress the underlying spaces.

Notation Switch: $P \leftrightarrow P$

Comment: Transition kernels from X to Y should be thought of as random transformations of an X -valued random variable $X \sim P_X$ into a Y -valued random variable $Y \sim P_Y$.

Visually,



Indeed,

$$P_Y(B) \triangleq \mathbb{E}_{P_x}[K(B|x)] \quad B \in \mathcal{G}$$

$$= \int_X K(B|x) dP_x(x)$$

Proposition: P_Y as defined above is a probability measure on (Y, \mathcal{G})

Proof

$$\begin{aligned} (i) \quad P_Y(Y) &= \int_X K(Y|x) dP_x(x) \\ &\quad \text{I for } \sum_{x \in X} \\ &= \int_X dP_x(x) = 1 \end{aligned}$$

(ii)

$$P_Y \left(\bigcup_{n=1}^{\infty} B_n \right) = \int_X K \left(\bigcup_{n=1}^{\infty} B_n \mid z \right) dP_X(z)$$

Upper bounded
by 1

$$= \int_X \sum_{n=1}^{\infty} K(B_n \mid z) dP_X(z)$$

(Fubini's Theorem)

$$= \sum_{n=1}^{\infty} \int_X K(B_n \mid z) dP_X(z)$$

$$= \sum_{n=1}^{\infty} P_Y(B_n)$$

Conditional Expectation

$$\begin{aligned} h(x) &= \mathbb{E}[Y \mid X=x] \\ h(X) &= \mathbb{E}[Y \mid X] \end{aligned} \quad \left. \begin{array}{l} \text{what is introduced} \\ \text{in other courses} \end{array} \right\}$$

Definition - Conditional Expectation

Let $K(\cdot \mid \cdot)$ be a transition kernel from X to Y .

Define

$$\textcircled{1} \quad \mathbb{E}[Y \mid X=x] \triangleq \int_Y y dK(y \mid x)$$

$$\textcircled{2} \quad \mathbb{E}[Y \mid X] \triangleq \int_Y y dK(y \mid X)$$

Theorem - Law of Total Expectation

$$E[Y] = E_{P_X}[E[Y|X]]$$