LECTURE 1

www.cs.hunter.cuny.edu/~felisav/StochasticProcesses/welcome.html

1. Go through outline and plan for course

Motivation for course: modeling and simulation (real life problems)

CHAPTER 1: PROBABILITY SPACES, RANDOM VARIABLES

In order to study complex dynamical systems that have uncertainty, we need to build appropriate models.

Model Components:

(a). EVENTS: need to introduce what we want to talk about and describe in our model: outcomes or states of nature

(b) INFORMATION STRUCTURE: how outcomes and dynamics are related

(c) RANDOM VARIABLES: measurements from the observation of the system

(d) PROBABILITY: assigning likelihoods and computing likelihoods of complex events.

(a) (b) INFORMATION STRUCTURE

The model starts by defining a sample space Ω. Elements $\omega \in \Omega$ represent "outcomes" (experimental setting in terminology), such as the result of market indices, the result of throwing dice, etc.

Experiment: throughout this chapter we will be using a thought experiment where two coins are tossed in separate rooms

nickel	dime
H_n, T_n	H_d, T_d

$$\Omega = \{(H_n, H_d), (H_n, T_d), (T_n, H_d), (T_n, T_d)\} = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

The information structure is built so that ~~we~~ it contains all the relevant events and observables in our model. A consistent model (logically) should satisfy being a σ-algebra:

Def: A collection $\mathcal{F}$ of ~~sets~~ subsets of Ω is called a σ-algebra if it satisfies:

(a) $\phi \in \mathcal{F}$ $\Omega \in \mathcal{F}$

(b) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(c) If $A_1, \dots A_n, \dots \in \mathcal{F} \Rightarrow \bigcup_{n \geq 1} A_n \in \mathcal{F}$

Def: A subset of Ω is called an event.

Def: A space $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-algebra on Ω, is called a probability space.

Example: the result from the coin-tossing triggers a response that is different if the coins have different outcomes, as follows: a light turns red if both are heads, blue if both are tails, and green if they result in different outcomes.

Describing the "observable" events requires ~~three~~ different models for different perspectives.

Let $\mathcal{F}$ be the collection of all possible events of Ω: $\{\omega_i\} \in \mathcal{F}$ $i = 1, \dots 4$, all unions and complements also, to form a σ-algebra. (Page 18 TK).

Consider now:

1) The collection of all possible events as seen by player 1 (n)

2) " " " " " " " 2 (d)

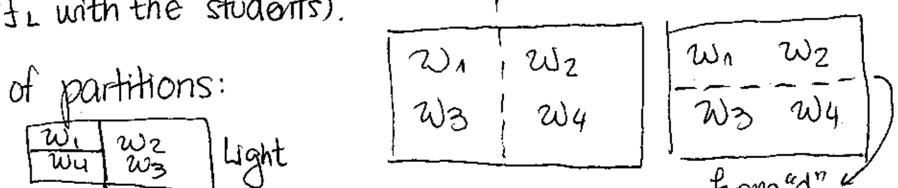
3) The collection of all possible events described by the light.

$\mathcal{F}_d = \{\phi, \Omega, \{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}\}$

$\mathcal{F}_n = \{\phi, \Omega, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\}$ "view from d"

(build $\mathcal{F}_L$ with the students).

Concept of partitions:



Def: A family of σ-algebras $\mathbb{F} = \{\mathcal{F}_n\}_{n=1}^{\infty}$ is a filtration on $(\Omega, \mathcal{F})$ if each $\mathcal{F}_n \subset \mathcal{F}$ is a σ-algebra and $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.

Increasing "levels" of detail in the information structure.

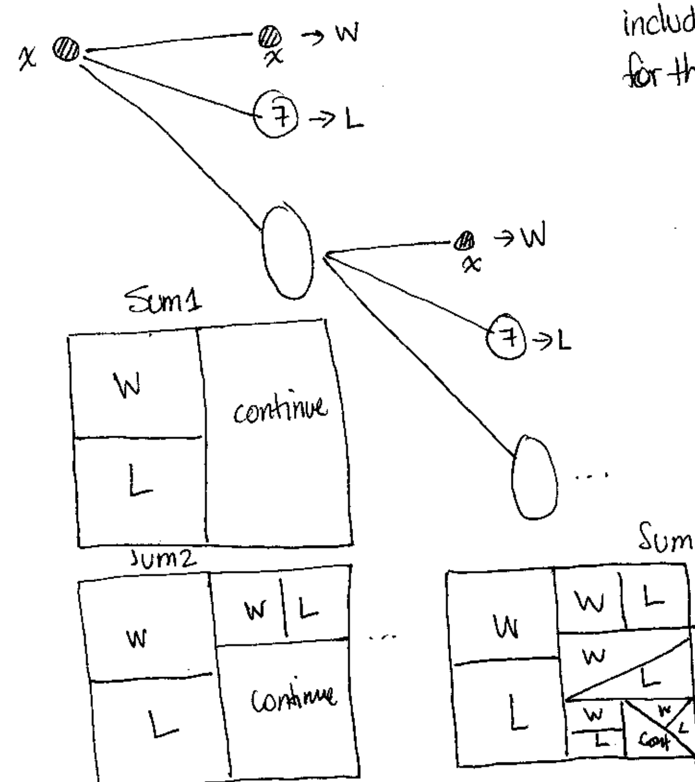
Example: game of craps p. 65 TK

Sum 1 Sum 2 Sum3

2, 3, 12 → L

7, 11 → W

The concept of "time" evolving is captured by increasing σ-algebras: every event contained in "the past" is also included in the structure for the "future".



Finer partitions of Ω: more detail

Backtrack: events up to "Sum6" $\mathcal{F}_6$ distinguish the history of the trajectory.

REMARK:
In the case of a FINITE set Ω, σ-algebras are related to partitions of Ω, according to the level of detail that they describe, and given a σ-algebra $\mathcal{F}$ on Ω, it is always possible to define a partition $\{A_1, \dots A_m\}$ of Ω such that $\mathcal{F}$ contains each $\{A_i\}$, as well as all unions and complements.. In our example of coin tossing:

$\mathcal{F}_d$ is related to the partition $A_1 = \{\omega_1, \omega_3\}$
$A_2 = \{\omega_2, \omega_4\}$

$\mathcal{F}_n$ " " " " $A_1 = \{\omega_1, \omega_2\}$
$A_2 = \{\omega_3, \omega_4\}$

Exercise: Describe the successive partitions of Ω in the game of craps (and notice that Ω is countable here).

As n increases, the corresponding partitions associated with a filtration $\{\mathcal{F}_n\}$ are increasing, becoming finer.

REMARK: A "probability space" $(\Omega, \mathcal{F})$ does not require a "prob".

RANDOM VARIABLES $(\Omega, \mathcal{F})$ given

Ask students to express what their understanding is of RVs "variables that take values by chance", "uncertainty", etc.

Def: (TK p17) A function $X : \Omega \to \mathbb{R}$ is measurable w.r.t $\mathcal{F}$ if for any real numbers $a, b \in \mathbb{R}$ the event $\{\omega : a < X(\omega) \le b\} \in \mathcal{F}$.

$\boxed{\text{In particular, } \{\omega : X(\omega) \le x\} \in \mathcal{F} \ \forall x \in \mathbb{R}}$

Def: Let $(\Omega, \mathcal{F})$ be a probability space. X is called a random variable if $X : \Omega \to \mathbb{R}$ is a real-valued measurable function w.r.t. $\mathcal{F}$ (or "$\mathcal{F}$-mbl" for short).

Def: The σ-algebra generated by a random variable X on $(\Omega, \mathcal{F})$, denoted $\sigma(X)$ is the smallest σ-algebra w.r.t which X is measurable.

A "construction" approach for countable Ω: partition the event set Ω as follows:

$$A_j = \{\omega : X(\omega) = x_j\} \quad j = 1, 2, \dots$$

Then build $\sigma(X)$ as the smallest σ-algebra containing each $\{A_j\}$.

Terminology: X a r.v. on $(\Omega, \mathcal{F})$. Suppose that $\sigma(X) \subset \mathcal{G} \subset \mathcal{F}$ for some σ-algebra $\mathcal{G}$. Then X is also $\mathcal{G}$-mbl (exercise for students).

Example: nickel and dime, page 18 TK:
X_d, X_n are rv's on $(\Omega, \mathcal{F})$

Have students work out if X_d is $\mathcal{F}_n$-mbl?

note that $\{X_d \le 0\} = \{\omega_1, \omega_3\} \notin \mathcal{F}_n$
$\{X_d > 0\} = \{\omega_2, \omega_4\} \notin \mathcal{F}_n$

Is X_n $\mathcal{F}_2$-mbl? $\longrightarrow$ have students work this out.

<u>Def</u>: Let (X,Y) be two random variables defined on a common probability space $(\Omega,\mathcal{F})$. Then $\sigma(X,Y)$ is the smallest σ-algebra w.r.t. which both X and Y are measurable.

<u>Example</u>: Build $\sigma(X,Y)$ for example on p. 18 TK (nickel and dime) and compare with $\sigma(Z)$.

By construction $\sigma(X) \subset \sigma(X,Y)$ and $\sigma(Y) \subset \sigma(X,Y)$.

<u>Definition</u>: Given $(\Omega,\mathcal{F})$ and a filtration $\mathbb{F} = \{\mathcal{F}_t\}$, $\mathcal{F}_t \subset \mathcal{F}$ $\forall t \in T$, a <u>stochastic process</u> $\{X_t ; t \in T\}$ is a collection of random variables such that for each t, X_t is $\mathcal{F}_t$-mbl.

The particular case $\mathcal{F}_t = \sigma(X_s ; s \le t)$ is called the <u>natural filtration</u> of the process.

REMARK: The notion of "time" dynamics (the arrow of time) is captured by the filtration or "history" of the process, as opposed to a "field" where knowledge in one direction does not necessarily pass to other directions.

<u>Example</u>: Visualize the game of craps as a stochastic process.

<u>Def</u>: A <u>random stopping time</u> adapted to the filtration $\{\mathcal{F}_t, t \in T\}$ on $(\Omega,\mathcal{F})$ is a random variable τ on $(\Omega,\mathcal{F})$ that satisfies:

$$\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t \quad \forall t \in T.$$

The σ-algebra $\mathcal{F}_t$ contains all events A such that $\{A \cap \tau(\omega) \le t\} \in \mathcal{F}_t$.

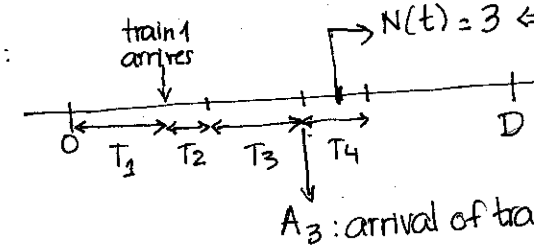
<u>Example</u>: subway station opens at time 0 and closes at the end of the day D.

$\{T_i\}$: times between consecutive arrivals

$A_n = \sum_{i=1}^{n} T_i$ arrival time of n-th train

$N(t) = \#$ trains up to time t $= \max(n : A_n \le t)$

Picture:



Let τ be the number of trains in one day, that is:

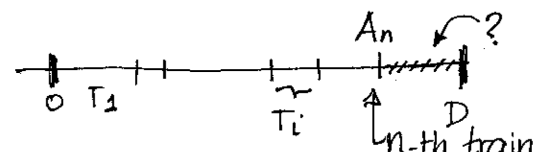
$$\boxed{\tau = N(D)}$$

Let $\mathcal{F}_n = \sigma(T_1, \dots, T_n)$ and consider the filtration $\{\mathcal{F}_n\}$.

The process $\{A_n ; n \ge 1\}$ is adapted to $\{\mathcal{F}_n\}$.

⇨ Is τ a stopping time w.r.t. $\{\mathcal{F}_n\}$? (Let them think)

Notice that the events $\{\tau = n\} \Leftrightarrow \{\omega : A_n \le D < \underbrace{A_{n+1}}_{= A_n + T_{n+1}}\}$



whether $\tau = n$ depends on the length of T_{n+1}, which is not mbl-$\mathcal{F}_n$ $\Rightarrow \{\tau = n\} \notin \mathcal{F}_n$. Information up to A_n is not enough to know that the last train of the day was number n.

Is $\tau + 1$ a stopping time w.r.t. $\{\mathcal{F}_n\}$?

LECTURE 2: PROBABILITY

~~(a) Probability measures~~

~~(b) Conditional Probability~~

~~(c) Expectations and Conditional Expectations~~

(a) Probability measures and Distributions

Mathematical model : $(\Omega, \mathcal{F})$ is the underlying space and information structure $\mathcal{F}$ containing all the possible events in the model.

Probabilities reflect our notion of "likelihood" for the events in $\mathcal{F}$. The model is constructive : adding the "odds" or likelihoods for disjoint events.

Def: A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a set function $\mathbb{P} : \mathcal{F} \to [0,1]$ satisfying:

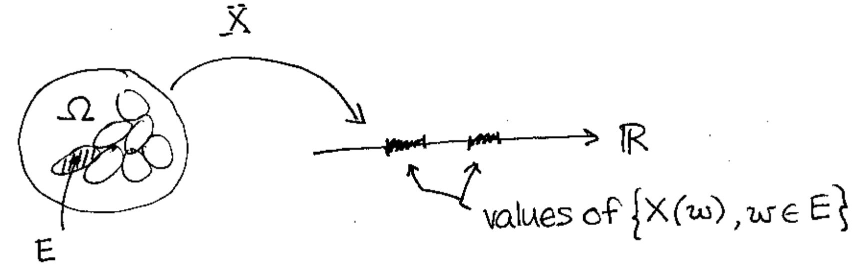
i). $\mathbb{P}(\phi) = 0$, $\mathbb{P}(E) \geq 0 \ \forall\ E \subset \mathcal{F}$

ii). For any countable collection $\{A_1, \ldots A_n\} \in \mathcal{F}$ of disjoint events $\mathbb{P}(\bigcup_{i \geq 1} A_i) = \sum_{i \geq 1} \mathbb{P}(A_i)$.

iii). $\mathbb{P}(\Omega) = 1$.

[Example 1.3 p.5 Ross]

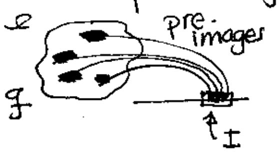
Let X be a random variable on $(\Omega, \mathcal{F})$. Because $X : \Omega \to \mathbb{R}$, this mapping induces a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel sets in $\mathbb{R}$ (the σ-algebra generated by the intervals in $\mathbb{R}$).



Therefore, we can associate a likelihood to the numerical outcome of X. For any interval $I \subset \mathbb{R}$, we can associate a probability to the "event":

$$\{X \in I\} \equiv \{\omega : X(\omega) \in I\} \subset \mathcal{F}$$

(why?)



Def: The probability distribution of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is $F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}\{\omega : X(\omega) \leq x\}$.

Notation: $\mathbb{P}_X((a,b]) = \int_a^b dF_X(x)$.

[Start out with dice or a Bernoulli : "what is Ω?"]

Theorem: Skorohod Representation

Given a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ there exists a function $\tilde{X} : [0,1] \to \mathbb{R}$ on the canonical space $([0,1], \mathcal{B}[0,1], \tilde{\mathbb{P}})$, where $\tilde{\mathbb{P}}$ is the Uniform distribution (or Lebesgue measure) such that $X \stackrel{d}{=} \tilde{X}$ (equal in distribution).

[proof will be done when we see inverse function method].

$\to$ Example $U \sim U(0,1)$ and $1-U \sim U(0,1)$.

Reading material from TK+Ross. Look @ examples.

Def: Two rv's X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$ are :

- equal almost surely $X = Y$ a.s. (or w.p.1) if $\mathbb{P}(\omega : X(\omega) = Y(\omega)) = 1$
- equal in dist. $X \stackrel{d}{=} Y$ if $\forall x \in \mathbb{R}$ $\mathbb{P}(X \leq x) = \mathbb{P}(Y \leq x)$

(a) Expectations

~~Def~~ Consider a countable sample space Ω with information structure $\mathcal{F}$. Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and define, for every possible value x_i of X:

$$A_i = \{\omega : X(\omega) = x_i\}.$$

Clearly $\{A_i, i \geq 1\}$ is a partition of Ω, and $\sigma(X)$ is the σ-algebra generated by this partition.

Def: The expectation of a rv X is the integral:

$$\mathbb{E}X = \sum_{i \geq 1} x_i \mathbb{P}(A_i) = \int x \, P_X(dx)$$

and it is well defined whenever $|\mathbb{E}X| < \infty$.

Thus on the case that $\mathcal{F}$ is a σ-algebra generated by a countable partition $\{A_1, A_2, \dots\}$ of Ω, any $\mathcal{F}$-mbl function has an expectation (if finite) which is a weighted sum of the atomic values of X.

If X is a continuous rv, with distribution $F_X(\cdot)$ then its density $f_X = \frac{dF}{dx}$ is well defined and

$$\mathbb{E}X = \int x f_X(x) dx.$$

(b) Conditioning

~~PP/8/8 handout~~

Example: $\Omega = \{\omega_1, \dots \omega_8\}$ represents 8 individuals. Prob space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{P}(\{\omega_i\}) = 1/8$ (uniform). Salaries:

$$\underbrace{\begin{matrix} X(\omega_1) = 25{,}000 \\ X(\omega_2) = 35{,}000 \end{matrix}}_{\text{low } (A_1)} \qquad \underbrace{\begin{matrix} X(\omega_3) = 50{,}000 \\ X(\omega_4) = 60{,}000 \\ X(\omega_5) = 70{,}000 \\ X(\omega_6) = 80{,}000 \end{matrix}}_{\text{medium } (A_2)} \qquad \underbrace{\begin{matrix} X(\omega_7) = 85{,}000 \\ X(\omega_8) = 95{,}000 \end{matrix}}_{\text{high } (A_3)}$$

Let $Y : \Omega \to \mathbb{R}$ denote the salary bracket

$$Y(\omega) = i \quad \text{if } \omega \in A_i$$

Conditional expectations $\mathbb{E}(X | Y = y)$ are averages within the given bracket:

$$\mathbb{E}(X | Y = 1) = 30{,}000$$
$$\mathbb{E}(X | Y = 2) = 65{,}000$$
$$\mathbb{E}(X | Y = 3) = 90{,}000$$

This means that the conditional expectation $\mathbb{E}(X|Y)$ is a random variable:

$$Z(\omega) = \begin{cases} 30{,}000 & \omega \in A_1 \\ 65{,}000 & \omega \in A_2 \\ 90{,}000 & \omega \in A_3 \end{cases}$$

Remark: If $Y' \neq Y$ has constant values on A_1, A_2 and A_3 $\Rightarrow \sigma(Y) = \sigma(Y')$. Is $\mathbb{E}(X | Y') = Z$ a.s.?
what enters in the averaging is the information structure generated by Y, rather than the actual values of Y.

Instead of defining $\mathbb{E}(X|Y)$ with two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, it then suffices to define $Z = \mathbb{E}(X|\mathcal{G})$, where $\mathcal{G}$ is a σ-algebra, $\mathcal{G} \subset \mathcal{F}$.

In the example:

$$\mathcal{G} = \sigma(Y) = \{\phi, \Omega, A_1, A_2, A_3, \{A_1, A_2\}, \{A_1, A_3\}, \{A_2, A_3\}\}$$

and $Z(\omega)$ is constant along the partition $\{A_1, A_2, A_3\}$ of Ω
$\Leftrightarrow$ Z is $\mathcal{G}$-mbl.

Def : Let X be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}|X| < \infty$ and let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. The conditional expectation $\mathbb{E}(X|\mathcal{G})$ is a random variable satisfying:

(a) Z is $\mathcal{G}$-mbl

(b) For all $A \in \mathcal{G}$, $\mathbb{E}(Z \mathbb{1}_A) = \mathbb{E}(X \mathbb{1}_A)$,

where $\mathbb{1}_A$ is the indicator function of the event A:

$$\mathbb{1}_A(\omega) = \begin{cases} 0 & \omega \notin A \\ 1 & \omega \in A \end{cases}$$

In particular, when $\mathcal{G} = \sigma(A_1, \ldots A_n)$ is the σ-algebra generated by a finite partition $\{A_1, \ldots A_n\}$ of Ω, then

$$Z(\omega) = \mathbb{E}(X|\mathcal{G})(\omega) = \sum_{i=1}^{n} \frac{\mathbb{E}(X \mathbb{1}_{A_i})}{\mathbb{P}(A_i)} \mathbb{1}_{A_i}(\omega)$$

Summary of Properties of Conditional Expectations

Theorem. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, random variables X and Y on $(\Omega, \mathcal{F}, \mathbb{P})$, and sub-$\sigma$-algebras $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2$ of $\mathcal{F}$.

A. If X is $\mathcal{G}$-mbl then $\mathbb{E}(X|\mathcal{G}) = X$ a.s.

B. $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F} \Rightarrow$

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G}_1)|\mathcal{G}_2) = \mathbb{E}(\mathbb{E}(X|\mathcal{G}_2)|\mathcal{G}_1) = \mathbb{E}(X|\mathcal{G}_1)$$

C. $\mathbb{E}(X|\{\phi, \Omega\}) = \mathbb{E}X$

D. For any function h with $\mathbb{E}(h(X)) < \infty$

$$\mathbb{E}(\mathbb{E}(h(X)|\mathcal{G})) = \mathbb{E}(h(X)).$$

E. If Y is $\mathcal{G}$-mbl and both $\mathbb{E}X$ and $\mathbb{E}(XY)$ exist then $\mathbb{E}(XY|\mathcal{G}) = Y\,\mathbb{E}(X|\mathcal{G})$.

F. (Jensen's inequality) : if h is a convex function and $\mathbb{E}|h(X)| < \infty$ then $h(\mathbb{E}(X|\mathcal{G})) \le \mathbb{E}(h(X)|\mathcal{G})$ a

(see TK p. 61)

~~Let X be a r.v. on (Ω, F, P). Then the~~

For any $A \subset \mathcal{F}$, the rv $\mathbb{1}_A : \Omega \to \{0,1\}$ is well defined and $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$. Therefore we can also define conditional probabilities w.r.t. σ-algebras $\mathbb{P}(A|\mathcal{G})$, $\mathcal{G} \subset \mathcal{F}$, for any $A \in \mathcal{F}$.

Examples DD: file compression (ask students about "decompressing").

(c) <u>Independence of random variables</u>

<u>Def</u>: Given $(\Omega, \mathcal{F}, \mathbb{P})$ and a set of sub σ-algebras $\mathcal{G}_1, \dots \mathcal{G}_n \subset \mathcal{F}$, we say that $\mathcal{G}_1, \dots \mathcal{G}_n$ are independent if for all sets $G_i \in \mathcal{G}_i$, $i = 1, \dots n$

$$\mathbb{P}(G_1 \cap \dots \cap G_n) = \prod_{i=1}^{n} \mathbb{P}(G_i).$$

<u>Def</u>: The random variables $X_1, \dots X_n$ on $(\Omega, \mathcal{F}, \mathbb{P})$ are independent if their generated σ-algebras $\sigma(X_i), i=1,..n$ are independent.

In particular, $X_1, \dots X_n$ are independent $\iff$

$$\mathbb{P}(X_1 \in A_1, \dots X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i) \qquad (*)$$

for any set of intervals $(A_i \subset \mathbb{R};\ i = 1, \dots n)$.

<u>Notation</u>: write $X \perp\!\!\!\perp Y$ when X and Y are independent.

<u>Proposition</u>: Let X, Y be two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. If X and Y are independent, then $\mathbb{E}(X \mid \sigma(Y)) = \mathbb{E}X$.

<u>Remark</u>: Because, for any event $A \subset \mathcal{F}$ the rv $\mathbb{1}_{(A)}$ is well defined, then independence of events is defined using $(*)$: $A, B \in \mathcal{F}$ are independent $\iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

<u>Computing Expectations by conditioning</u>: see examples: p.64 TK Dice game craps; p.114 Ross quicksort; p.126 Ross best price game (Recursion $\iff$ renewal arguments)

Key formulas for computations: X, Y on $(\Omega, \mathcal{F}, \mathbb{P})$

$$\mathbb{E}(X) = \sum_{i \geq 1} \mathbb{E}(X \mid Y = y_i)\,\mathbb{P}(Y = y_i)$$

when $Y \in \{y_i, i \geq 1\}$ is countable (discrete rv).

$$\mathbb{E}(X) = \int \mathbb{E}(X \mid Y = y)\, f_Y(y)\, dy,$$

when f_Y is the density of cts rv Y.

<u>Reading material</u>: particularly 3.6 Ross in addition to all examples in TK + Ross.

<u>Def</u>: The variance of a random variable X on $(\Omega, \mathcal{F})$ is:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

<u>Proposition</u>: Given a r.v. X on $(\Omega, \mathcal{F}, \mathbb{P})$ and a σ-algebra $\mathcal{G} \subset \mathcal{F}$,

$$\boxed{\mathrm{Var}(X) = \mathbb{E}\big(\mathrm{Var}(X \mid \mathcal{G})\big) + \mathrm{Var}\big(\mathbb{E}(X \mid \mathcal{G})\big)}$$

Notice that if we define $Z = \mathbb{E}(X \mid \mathcal{G})$, then the conditional variance is the random variable:

$$\mathrm{Var}(X \mid \mathcal{G}) = \mathbb{E}(X^2 \mid \mathcal{G}) - Z^2$$

which is $\mathcal{G}$-mbl. (p.119 Ross)

[Convergence of random variables, pp. 5-7.13 in handout notes]

If time permits, work out game of craps solution to illustrate computing by conditioning & recursion: p. 65-66 TK.

(d) Radon-Nikodym Theorem

Def: Given a probability space $(\Omega, \mathcal{F})$ a function $\mu : \mathcal{F} \to \mathbb{R}$ is called a measure if

- $\mu(E) \geq 0$ for all $E \subset \mathcal{F}$
- $\mu(\phi) = 0$.
- for any countable collection $\{E_1, E_2, \ldots\} \subset \mathcal{F}$,

$$\mu\left(\bigcup_{i \geq 1} E_i\right) = \sum_{i \geq 1} \mu(E_i).$$

In particular, if $\mu(\Omega) = 1$ then μ is a probability measure.

When the function μ satisfies $\mu(\phi) = 0$ and is countably additive but not necessarily non-negative, it is called a signed-measure.

Terminology: $(\Omega, \mathcal{F})$ is called a measurable space, while $(\Omega, \mathcal{F}, \mu)$ is called a measure space.

When $|\mu(\Omega)| < \infty$, the measure is called sigma finite (or just finite).

Def: Let μ, ν be two finite measures on $(\Omega, \mathcal{F})$. Then $\mu \ll \nu$ (μ is absolutely continuous w.r.t. ν) if $\forall A \in \mathcal{F}$ $\nu(A) = 0 \Rightarrow \mu(A) = 0$.

- If $\nu \ll \mu$ and $\mu \ll \nu$ then the two measures are ~~called~~ said to be equivalent.
- If a family of measures $\{\mu_n\}$ on $(\Omega, \mathcal{F})$ are all absolutely continuous w.r.t. ν, we say that the family is dominated by ν.

Consider $(\Omega, \mathcal{F})$ a given measurable space. Let $L : \Omega \to \mathbb{R}^+$ be a ~~rv~~ function satisfying:

(a) L is measurable w.r.t. $\mathcal{F}$

(b) Given a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$, L satisfies

$$\mathbb{E}_P(L) = \int_\Omega L(\omega) \mathbb{P}(d\omega) = 1.$$

(c). L is non-negative for every $\omega \in \Omega$.

We will now use this random variable to construct another probability measure on $(\Omega, \mathcal{F})$ as follows

For each $A \in \mathcal{F}$, define

$$\mathbb{Q}(A) = \int_A L(\omega) \mathbb{P}(d\omega)$$

Exercise: verify that $\mathbb{Q} : \Omega \to [0,1]$ is a probability measure and that it satisfies $\mathbb{Q} \ll \mathbb{P}$.

Note: whether $\mathbb{P} \ll \mathbb{Q}$ depends on the null sets

$$N = \{\omega : L(\omega) = 0\}.$$

Indeed, if $A \subset \mathcal{F}$ is such that $\mathbb{P}(A) > 0$ but $A \subset N$, then $\mathbb{Q}(A) = 0$, in which case $\mathbb{P} \ll \mathbb{Q}$ does NOT hold.

The following theorem states the converse result to the construction of $\mathbb{Q}$ from a rv L, and it is an important basis for a number of techniques involving changes of measure (math finance, simulation, etc).

<u>Theorem</u> (Radon-Nikodym): Two sigma-finite measures on $(\Omega,\mathcal{F})$ satisfy $\mu \ll \nu$ if, and only if, there exists a non-negative random variable L on $(\Omega,\mathcal{F})$ such that for every $A \subset \mathcal{F}$

$$\mu(A) = \int_A L(\omega)\,\nu(d\omega) = \mathbb{E}_\nu(L\,\mathbb{1}_A).$$

The random variable L is called the <u>Radon-Nikodym derivative of μ w.r.t. ν</u> and it is denoted $\left[\frac{d\mu}{d\nu}\right]$. It is also called the <u>ν-density of μ</u>.

<u>Remark</u>: In the particular case where ν is the Lebesgue measure on $\mathbb{R}$ and $\mathbb{E}(L)=1$, then μ is a probability measure and $\left[\frac{d\mu}{d\nu}\right]$ is the "usual" density function.

<u>Remark</u>: μ is a probability measure $\Leftrightarrow$ $\mathbb{E}_\nu(L)=1$, and it is not necessary that ν be a probability measure.

<u>Example</u>: Let X_p be a Bernoulli (p) random variable. Using the canonical representation,

$$X_p(u) = \mathbb{1}_{(u \geq p)} \qquad u \in [0,1]$$
$$\mathbb{P}(u \in A) = |A|,\ A \subset [0,1].$$

Although this discrete rv does not have a Lebesgue density, it has a density w.r.t. $Y \sim Ber(0.5)$. Indeed:

$$L(0) = \frac{1-p}{0.5}, \quad L(1) = \frac{p}{0.5}$$

Consider any function $h:\mathbb{R}\to\mathbb{R}$, then clearly:

$$\mathbb{E}_p(h(X)) = (1-p)h(0) + p\,h(1)$$

$$= \mathbb{E}_{0.5}\Big(h(X)L(X)\Big) = \frac{(1-p)}{0.5}h(0)\times 0.5 + \frac{p}{0.5}h(1)\times 0.5$$

then the family of rv's (or distributions of) X_p is dominated by the Bernoulli (0.5).

<u>Example</u>:
Rare events and Change of Measure.

Suppose that $A \subset \Omega$, and given $(\Omega,\mathcal{F},\mathbb{P})$, $A \subset \mathcal{F}$ is such that $\mathbb{P}(A) \ll 1$.

Let X be a random variable such that $X(\omega)=0$ $\omega \notin A$ (the support is A).

Define: $L(\omega) = (\mathbb{P}(A))\,\mathbb{1}_{(\omega \in A)} \leq 1$ a.s.

Then for any $B \subset \mathcal{G} \equiv \sigma(X)$,

$$\mathbb{P}(B) = \int_B L(\omega)\,\mathbb{Q}(d\omega)$$

where $\mathbb{Q}(B) = \mathbb{P}(B|A) = \frac{(B\cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(A)}$ for $B \subset \sigma(X) \subset A$.

This is an example where $\mathbb{Q}(A)=1$, so the rare event becomes certain under $\mathbb{Q}$, and $\mathbb{Q} \ll \mathbb{P}$.

Is $\mathbb{P} \ll A$? (NO)

R-N Theorem: $\mu = \mathbb{P}|_A$ is the restriction of $\mathbb{P}$ to A (defective probability): $\mu(B) = \mathbb{P}(B)\ \forall B \in \mathcal{G} \subset \mathcal{F}$, and $\mu(\Omega) = \mathbb{P}(A) < 1$.