REVISION OF PROBABILITY

To understand complex dynamical systems under uncertainty (randomness) we need an appropriate model.

Model Components:

(a). EVENTS: introduce what we wish to describe as possible outcomes (also called "states of nature")

(b). INFORMATION STRUCTURE: introduces the manner in which an observer can describe relationships between events.

(c). RANDOM VARIABLES/PROCESSES: numerical measurement or 'proxies' associated with the outcomes or observations.

(d). PROBABILITY: assigns likelihoods to possible events and computes likelihoods and distributions.

[NEXT: Convergence of RV's and Martingales]

(a) EVENTS: In mathematical modeling, we start by stating a sample or event space Ω.

Elements $\omega \in \Omega$ represent possible "outcomes" (experimental setting terminology) such as:

- result of market indices
- result of game competitions
- results of sales of a company at end of day

(b). INFORMATION STRUCTURE

Thought Experiment: Two coins are tossed in separate rooms

nikel	dime
H_n, T_n	H_d, T_d

BB

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

The information structure is built to contain all relevant observables in our model. A logically consistent model should satisfy the properties of a σ-algebra.

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

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

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

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

Def: a subset E of Ω, $E \subset \Omega$ is called an event.

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

Example: Player 1 (n), Player 2 (d), outside light: red if both heads, blue if both tails, green otherwise.

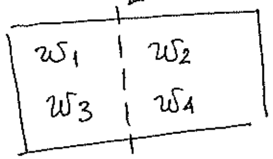
How would n, d, or ℓ describe the possible observables according to their own perspective?

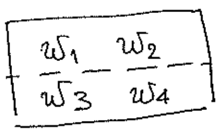
BB

$\mathcal{F}_d = \{ \phi, \Omega, \{\omega_1,\omega_3\}, \{\omega_2,\omega_4\} \}$ (build $\mathcal{F}_\ell$ with students)

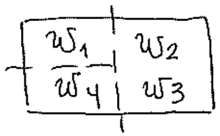
$\mathcal{F}_n = \{ \phi, \Omega, \{\omega_1,\omega_2\}, \{\omega_3,\omega_4\} \}$

view from d





view from n



light

Concept of partitions (when Ω is finite) and power sets.

Def: A family of σ-algebras $\mathbb{F} = \{\mathcal{F}_n\}_{n=1}^{\infty}$ is called 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 information structure: as time evolves, information is not lost. Concept of time is captured in filtration: events observable in past are also included in the future.

Example: Game of Craps.
Trow 2 dice and consider the sum of results (p. 65 TK)

Sum 1 | Sum 2 | Sum 3

2, 3, 12 → L
7, 11 → W
x → continue → x → W
7 → L
others → continue → x → W
7 → L
others → continue
...

The consecutive information structures are more detailed.

Remark: Finite $\Omega \Rightarrow \sigma$-algebras are related to partitions of Ω that describe the level of detail of observables in that model. Given a σ-algebra on finite Ω it is always possible to describe the corresponding partition $\{A_1, \dots, A_m\}$ of Ω such that $\mathcal{F}$ contains each A_i, their unions and intersections. Example:

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

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

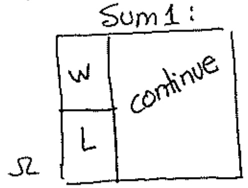
Exercise: describe successive partitions of Ω for game of craps

REMARK: A probability space $(\Omega, \mathcal{F})$ can be defined without the need to define a probability on it!

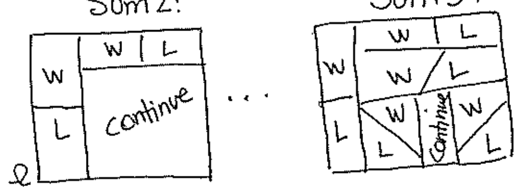
Example: Game of Craps

Visualization of consecutive partitions that will determine the filtration:

Sum 1:



Sum 2: ... Sum 5:



Backtrack: events described up to the σ-algebra at step 5 (Sum 5) $\mathcal{F}_5$ can distinguish the "history" of the process (game) up to 5 stages.

(c) RANDOM VARIABLES AND PROCESSES

Ask students to express what they understand of RV's.

Let $(\Omega, \mathcal{F})$ be a prob space.
Def: a function $X : \Omega \to \mathbb{R}$ is measurable with respect to $\mathcal{F}$ if $\forall\, a, b \in \mathbb{R}$ the event

$$\{\omega : a < X(\omega) \le b\} \in \mathcal{F}$$

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

Def: Let $(\Omega, \mathcal{F})$ be a probability space. X is called a random variable on $(\Omega, \mathcal{F})$ if $X : \Omega \to \mathbb{R}$ is a real-valued mbl 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 Ω as follows:

$$A_j = \{\omega : X(\omega) = \cancel{x}_j\}, \quad j = 1, 2, \ldots$$

for all possible values of X. Then build $\sigma(X)$ with A_j's, their unions and intersections.

Exercise: X a rv on $(\Omega, \mathcal{F})$. Suppose that $\sigma(X) \subset \mathcal{G} \subset \mathcal{F}$ for some σ-algebra $\mathcal{G}$. Then X is $\mathcal{G}$-mbl.

Example: Nickel and Dime

X_d = # Heads for dime

X_n = # Heads for nickel

$Z = X_d + X_n$ (how is this related to light?)

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

Note: $\{X_d \le 0\} = \{\omega_2, \omega_4\} \notin \mathcal{F}_n$

$\{X_d > 0\} = \{\omega_1, \omega_3\} \notin \mathcal{F}_n$

Is X_n $\mathcal{F}(Z)$-mbl? $\to$ have students work this out.

Def: Let X, Y be two rv's 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.

Example: build $\sigma(X, Y)$ for the nickel and dime example and compare with $\sigma(Z)$. [Left as exercise for students].

Def: Given $(\Omega, \mathcal{F})$ and a filtration $\mathbb{F} = \{\mathcal{F}_t, t \in T\}$ on $(\Omega, \mathcal{F})$, a stochastic process $\{X_t, t \in T\}$ is a collection of random variables such that for each $t \in T$, X_t is $\mathcal{F}_t$-mbl.

The special case $\mathcal{F}_t = \sigma(X_s;\, s \le t)$ is called the natural filtration of the process.

REMARK: The notion of "time" dynamics (the arrow of time) is captured by the filtration, or "history" of the process, so that information structure from the past is carried on to the future (not lost).

(d). PROBABILITY AND EXPECTATIONS

Mathematical model : $(\Omega, \mathcal{F})$ is the underlying probability space and the information structure $\mathcal{F}$ describes all possible events in the model. Probabilities associated to events reflect our notion of "likelihood" for events in $\mathcal{F}$. The model is constructive : adding the likelihoods of (disjoint) events.

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

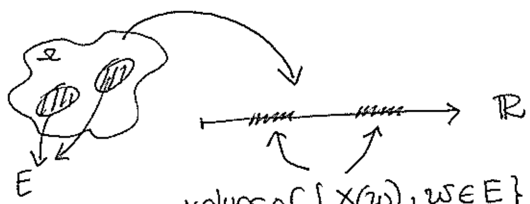
(i). $\mathbb{P}(\phi) = 0$, $\mathbb{P}(E) \geq 0 \quad \forall E \in \mathcal{F}$

(ii). For any countable collection $\{A_1, A_2, ..\}$, $A_n \in \mathcal{F}$ of disjoint events ($A_j \cap A_k = \phi$ if $i \neq k$)

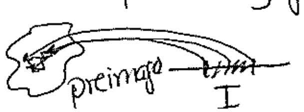
$$\mathbb{P}\Big(\bigcup_{n\geq 1} A_n\Big) = \sum_{n \geq 1} \mathbb{P}(A_n)$$

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

Let X be a rv on $(\Omega, \mathcal{F})$. Because X is a mapping $X : \Omega \to \mathbb{R}$, 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 numerical outcomes of X. For any interval $I \in \mathbb{R}$ we can associate the probability of the event : $\{\omega : X(\omega) \in I\} \subset \mathcal{F}$ (why?)



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

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}\{\omega : X(\omega) \leq x\}, \quad x \in \mathbb{R}.$$

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

We will now introduce the concept of expectations and conditional expectations using the case of countable Ω with information structure $\mathcal{F}$.

Let X be a r.v. on $(\Omega, \mathcal{F}, \mathbb{P})$ with countable state space S.

For each $x_i \in S$ define $A_i = \{\omega : X(\omega) = x_i\} \subset \mathcal{F}$.

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

Def The expectation of a rv X on $(\Omega, \mathcal{F}, \mathbb{P})$ is the integral

$$\mathbb{E}[X] = \sum_{i \geq 1} x_i \mathbb{P}(A_i) = \int_{\mathbb{R}} x \, \mathbb{P}_X(dx),$$

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

Thus when $\mathcal{F}$ is the power set of a countable partition $\{A_1, A_2, ..\}$ the expectation of any random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ is a weighted sum of the atomic values $x_i \in S$. If X is a continuous rv with distribution $F_X(\cdot)$ then its probability density $f_X = \frac{dF_X}{dx}$ is well defined and

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x) \, dx.$$

CONDITIONING

To introduce the notion of conditioning we will use a simple examples. $\Omega = \{\omega_1, \dots, \omega_8\}$ represents 8 individuals. Probability space: $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is the power set of Ω and $\mathbb{P}(\{\omega_i\}) = 1/8$ (the likelihoods can be changed, but this uniform model makes the calculations easier). Salaries:

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

Let $Y : \Omega \to \mathbb{R}$ label the salary bracket: $Y(\omega) = i$ if $\omega \in A_i$.

Conditional expectation: $\mathbb{E}[X \mid Y = y]$ is a number which is the average within the given bracket.

$\Rightarrow$ $\mathbb{E}[X \mid Y = 1] = 30{,}000$

$\mathbb{E}[X \mid Y = 2] = 65{,}000$

$\mathbb{E}[X \mid Y = 3] = 90{,}000$

$\boxed{\text{what is } \mathbb{E}[X \mid Y] \text{ ?}}$ without specifying the actual value of Y, this conditional expectation is a function of Ω (because it is a function of Y). Define the random variable:

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

Then $\mathbb{E}[X \mid Y] = Z$ is a random variable on $(\Omega, \mathcal{F})$.

Remark: If $Y' \neq Y$ has constant values on A_1, A_2 and A_3, say $Y'(\omega) = e^{-5}$ if $\omega \in A_1$, $Y'(\omega) = -\pi$ if $\omega \in A_2$, $Y'(\omega) = 42$ if $\omega \in A_3$, then $\mathbb{E}[X \mid Y']$ will be the same rv as Z: what enters in the averaging are the values of X and the information structure (partition) rather than the actual values of Y or Y'.

Instead of defining conditional expectation $\mathbb{E}[X \mid Y]$ as a function of Y, it then suffices to define it in terms of a conditioning σ-algebra $\mathcal{G}$, for $\mathcal{G} \subset \mathcal{F}$.

In this example $\mathcal{G} = \sigma(Y)$:

$$\mathcal{G} = \{\phi, \Omega, A_1, A_2, A_3, \{A_1, A_2\}, \{A_1, A_3\}, \{A_2, A_3\}\}$$

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

Def: Let X be a rv 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 $\mathbb{E}[X \mid \mathcal{G}]$ 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 of the event A:

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

For $\mathcal{G} = \sigma(A_1, \dots A_n) \Rightarrow$

$$Z(\omega) = \mathbb{E}[X \mid \mathcal{G}](\omega) = \sum_{k=1}^{n} \frac{\mathbb{E}[X \mathbb{1}_{A_i}]}{\mathbb{P}(A_i)} \mathbb{1}_{A_i}(\omega).$$

[properties in Appendix B]

Def: Given $(\Omega,\mathcal{F},\mathbb{P})$ and a set of sub σ-algebras $\mathcal{G}_1,\mathcal{G}_2,\dots\mathcal{G}_n\subset\mathcal{F}$, we say that $\mathcal{G}_1,\dots\mathcal{G}_n$ are independent if for all sets $G_i\in\mathcal{G}_i$, $i=1,\dots n$

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

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

In particular, if $X_1\dots X_n$ are independent then for any set of intervals $\{A_i\subset\mathbb{R}, i=1,\dots n\}$ $\mathbb{P}(X_1\in A_1,\dots,X_n\in A_n)=\prod_{i=1}^{n}\mathbb{P}(X_i\in A_i)$.

Notation: write $X\perp\!\!\!\perp Y$ when X and Y are independent.

Proposition: Let X,Y be two rv's on $(\Omega,\mathcal{F},\mathbb{P})$. If $X\perp\!\!\!\perp Y$ then

$$\mathbb{E}[X\mid\sigma(Y)]=\mathbb{E}[X].$$

Remark: For any event $A\in\mathcal{F}$ the indicator $\mathbb{1}_A$ is a rv, so use the def. above to define independence between events. $A\perp\!\!\!\perp B \Leftrightarrow$ $\mathbb{P}(A\cap B)=\mathbb{P}(A)\mathbb{P}(B)$.

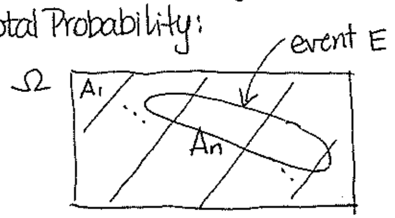
computing expectations by conditioning

Key formulas for computation and for proofs are an extension of the law of total probability (also related to Baye's formula).

Countable case: $Y\in S$ (countable) $\Rightarrow$ partition of $\Omega=\bigcup_{i\geq1}A_i$.

$A_i=\{w: Y(w)=y_i\}$, $y_i\in S$.

Total Probability:



$E=\bigcup_{n\geq1}(E\cap A_i)$

$\Rightarrow$ (disjoint sets)

$$\mathbb{P}(E)=\sum_{i\geq1}\mathbb{P}(E\cap A_i)$$
$$=\sum_{i}\mathbb{P}(E\mid A_i)\mathbb{P}(A_i)$$

In general: X a rv on $(\Omega,\mathcal{F},\mathbb{P})$ and $\{A_i\}$ a partition of Ω:

$$\mathbb{E}[X]=\sum_{i\geq1}\mathbb{E}[X\mid A_i]\mathbb{P}(A_i)$$

If Y has a density (continuous random variable) then

$$\mathbb{E}[X]=\int_{\mathbb{R}}\mathbb{E}[X\mid Y=y]f_Y(y)\,dy,$$

where f_Y is the density of Y.

CONVERGENCE OF RANDOM VARIABLES [Appendix B]

MARTINGALES [Appendix B]