$\boxed{\text{Probability Background}}$

- Probability spaces
- Random variables
- Expectation
- Basic facts from analysis

- Probability space $\equiv (\Omega, \mathcal{F}, \mathbb{P})$

sample space (Ω), σ-field (or σ-algebra) ($\mathcal{F}$), prob measure (or function) ($\mathbb{P}$)

- $\Omega \equiv$ set of all possible outcomes of an expt

- $\mathcal{F} \equiv$ collection of subsets of Ω with 3 prop -
  - i) $\Omega \in \mathcal{F}$
  - ii) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$
  - iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

- Eg - (Ω, ϕ), $(\Omega, \phi, A, \bar{A})$, 2^{Ω} for discrete Ω

- <u>Prop</u>

- Prop^n - For any collection G of subsets of Ω, $\exists$ a smallest σ-field $\sigma(G)$ that contains G

- Defn - For any metric space Ω (eg. $\mathbb{R}^n$), let $\mathcal{O}$ denote the collection of all open subsets of Ω. Then the Borel σ-field $\mathcal{B}(\Omega) = \sigma(\mathcal{O})$

- In particular, for $\mathbb{R}$, let $\mathcal{I} = \{(-\infty, a] \mid a \in \mathbb{R}\}$ be the set of closed intervals. Then

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I})$$

i.e., $\mathcal{B}(\mathbb{R})$ is the smallest σ-field containing all closed intervals (works also for open intervals)

- Note: This is a non-constructive definition. However, one can construct sets which are not in $\mathcal{B}(\mathbb{R})$!
(See wikipedia → Vitali set)

- $\mathbb{P} : \mathcal{F} \to [0,1]$ is a probability measure on σ-field $\mathcal{F}$ if
  i) $\mathbb{P}[\Omega] = 1$
  ii) $\mathbb{P}[A] \in [0,1] \ \forall A \in \mathcal{F}$
  iii) For $A_1, A_2, \ldots \in \mathcal{F}$ s.t. $A_i \cap A_j = \emptyset \ \forall i,j$

incompatible/disjoint/mutually exclusive

$$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$$

(KOLMOGOROV'S AXIOMS)

- The above axioms imply the foll^g

i) $\mathbb{P}[\phi] = 0$

ii) If $A \subseteq B$, then $\mathbb{P}[A] \leq \mathbb{P}[B]$

iii) $\mathbb{P}[\bar{A}] = 1 - \mathbb{P}[A]$

iv) For <u>any</u> $A_1, A_2, \ldots \in \mathcal{F}$

(Union bound) $$\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i]$$

v) Given $A_1, A_2, \ldots \in \mathcal{F}$ s.t $\mathbb{P}[A_i] = 0 \; \forall i$,

then $\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = 0$ (0-measure or negligible sets)

vi) Given $A_1, A_2, \ldots \in \mathcal{F}$ s.t $A_i \subseteq A_{i+1} \; \forall i$

then $\mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] = \lim_{i \to \infty} \mathbb{P}[A_i]$

(Sequential continuity)

$\boxed{\text{Try proving these / read up}}$

- Random variable $\equiv$ 'measurable' fn on Ω

- A fn $f : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ is measurable if $\forall A \in \mathcal{B}$, we have $f^{-1}(A) \in \mathcal{F}$

- (In this course, we will mostly ignore measurability - however in more complex settings, one needs to be careful...)

- <u>Defn</u> - A random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a measurable fn $X : \Omega \longrightarrow \mathbb{R}$

Almost Sure

- We are interested in all events which are non-negligible (i.e., all A s.t $\mathbb{P}[A] > 0$).

- **Defn** - For given $(\Omega, \mathcal{F}, \mathbb{P})$, a property P holds almost surely if $\mathbb{P}[\{\omega \mid \omega \text{ satisfies } P\}] = 1$

- Similarly for r.v.s X, Y, we say $X = Y$ almost surely (or $X = Y$ a.s.) if

$$\mathbb{P}[\{\omega \mid X(\omega) \neq Y(\omega)\}] = 0$$

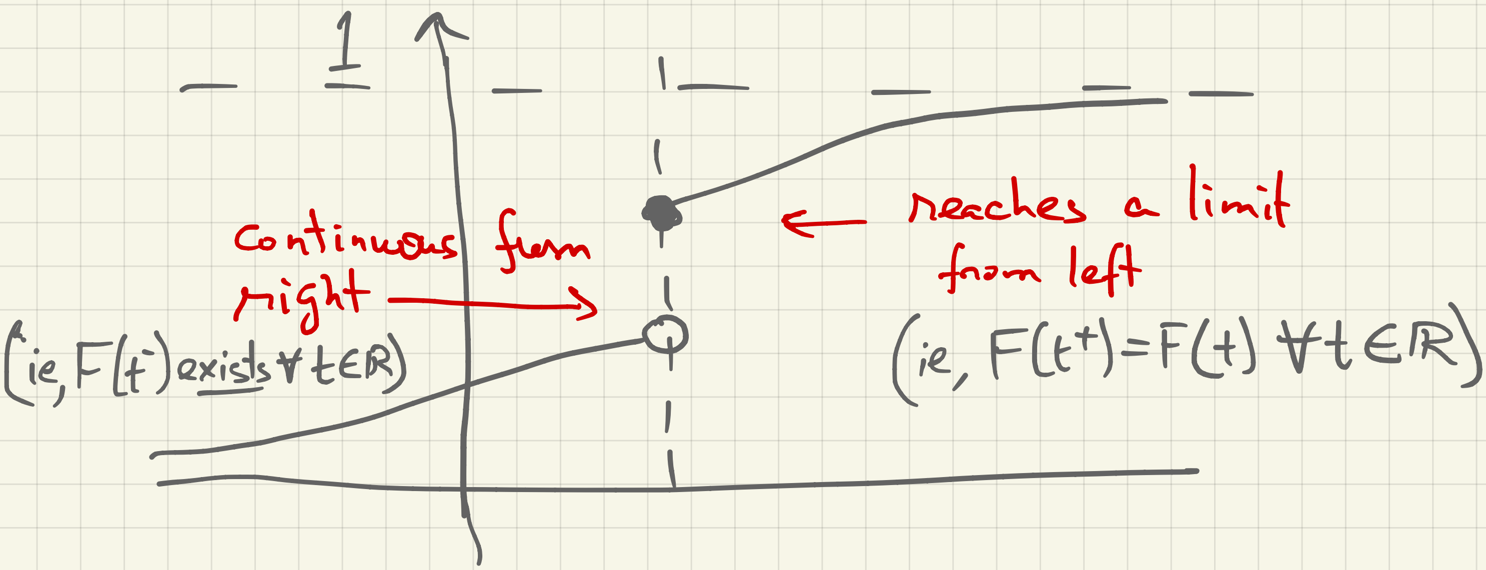
- The signature feature of a random variable X is its cumulative distribution fn $F: \mathbb{R} \to [0,1]$, $F(t) = \mathbb{P}[X \leq t]$
  - recall the Borel σ-field $\mathcal{B}(\mathbb{R})$ was constructed using only (closed) intervals $(-\infty, t]$
  - the $\leq$ is a convention; we could have as well defined it as $\mathbb{P}[X < t]$. This would correspond to constructing $\mathcal{B}(\mathbb{R})$ using open intervals ...
  - CDFs are awesome! All random variables have one.

Properties of CDFs

- $\lim_{t \to -\infty} F(t) = 0, \quad \lim_{t \to \infty} F(t) = 1$

- F is non-decreasing

- F is RCLL (Right continuous Left limits)



This is a consequence of $\leq$ convention

- Every non-decreasing, RCLL fn from 0 to 1 is a CDF for some X ...

- random variables can be discrete or continuous

Discrete	(Absolutely) Continuous
- $\exists\, x_1, x_2, \ldots \in \mathbb{R}$ s.t $\sum_{i=1}^{\infty} \mathbb{P}[X = x_i] = 1$	- $\exists$ fn $f: \mathbb{R} \to \mathbb{R}^+$ s.t. $F(t) = \int_{-\infty}^{t} f(x)\,dx$
- $p(x_i) = \mathbb{P}[X = x_i]$ probability mass fn (pmf)	probability density fn (pdf)
- $F(t) = \sum_{x_i \le t} p(x_i)$	- $\mathbb{P}[X = x] = 0 \quad \forall\, x \in \mathbb{R}$

- See Ch 2.1 of Brémaud (DiscProb) for examples of distributions

Discrete: Bernoulli(p), Binomial(n,p),
Geometric(p), Poisson(λ)
Multinomial$(m, p_1, p_2, \ldots, p_n) \equiv \begin{pmatrix} m \text{ balls in} \\ n \text{ bins} \end{pmatrix}$

Continuous: Uniform(a,b), Gaussian $\mathcal{N}(\mu, \sigma^2)$
Exponential(λ)

(for the last two, see Brémaud - Markov Chains)

- Random Vectors $\equiv$ Collection of random variables $(X_i)_{i\in I}$ on prob space $(\Omega, \mathcal{F}, \mathbb{P})$

- Joint distribution fn $F: \mathbb{R}^n \to [0,1]$

$$F(t_1, t_2, \dots, t_n) = \mathbb{P}[(X_1 \le t_1) \cap (X_2 \le t_2) \cap \dots \cap (X_n \le t_n)]$$

- Independence - 2 events $A, B \in \mathcal{F}$ are indep if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$
  - Different from mutually exclusive/disjoint

- For events $A_1, A_2, \dots, A_n$
  - $\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i]\,\mathbb{P}[A_j] \;\forall\, i,j$

    $\Rightarrow$ A_i's are Pairwise independent
  - $\mathbb{P}\left[\bigcap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} \mathbb{P}[A_i]$

    $\Rightarrow$ A_i's are mutually independent

- 2 random variables X, Y are independent if $\forall\, x, y \in \Omega$

$$\mathbb{P}[X=x, Y=y] = \mathbb{P}[X=x]\,\mathbb{P}[Y=y]$$

- r.v.s X, Y are said to be conditionally independent given Z if $\forall\, x, y, z$, we have

$$\mathbb{P}[X=x, Y=y \mid Z=z] = \mathbb{P}[X=x \mid Z=z]\,\mathbb{P}[Y=y \mid Z=z]$$

(see Brémaud for conditional probability)

- Notation - X, Y independent is written as $X \perp\!\!\!\perp Y$

Expectation

- We will focus first on discrete r.v.

<u>Defn</u> - For r.v. X taking values in countable set E, and function $g: E \to \mathbb{R}$ s.t. either g is non-negative or $\sum_{x \in E} |g(x)| \, p(x) < \infty$, 'integrable'

then
$$\mathbb{E}[g(X)] = \sum_{x \in E} \underbrace{p(x)}_{\text{pmf } \mathbb{P}[X=x]} g(x)$$

An important example

- For any $A \in \mathcal{F}$, $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]$,

where $\mathbb{1}_A \equiv$ indicator r.v $\left(\mathbb{1}_A = \begin{cases} 1 & \text{if } A \text{ true} \\ 0 & \text{o/w} \end{cases} \right)$

- Indicator r.v. are very useful for computations!!

- For continuous distributions

  - Suppose X has pdf f (ie, absolute continuous)

  then $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$

  for any g s.t $\int_{-\infty}^{\infty} |g(x)| f(x)\, dx < \infty$

  - More generally, X could be discrete + continuous

  - <u>FACT</u> - <u>Any</u> CDF F has only a countable number of jumps

  - Let $\{d_n\}_{n \geq 1}$ be the discontinuity points for a given CDF F

$$F_d(t) \triangleq \sum_{d_n \leq t} \left(F(d_n) - F(d_n^-) \right), \quad F_c(t) = F(t) - F_d(t)$$

'jumps' 'continuous part'

then $\mathbb{E}[g(X)] = \underbrace{\int_{-\infty}^{\infty} g(x)\, dF(x)}$

'the Lebesgue integral of g w.r.t measure F'

$$= \sum_{n=1}^{\infty} g(d_n)\left(F(d_n) - F(d_n^-)\right) + \int_{-\infty}^{\infty} g(x) f_c(x)\, dx$$

where $f_c(x)$ is a function such that $F_c(t) = \int_{-\infty}^{x} f_c(x)\, dx$

(i.e, $f_c(x) = \frac{d}{dx} F_c(x)$)

Properties of $\mathbb{E}[X]$

i) Linearity of Expectation

$$\mathbb{E}\left[\lambda_1 g_1(X) + \lambda_2 g_2(X)\right] = \lambda_1 \mathbb{E}[g_1(X)] + \lambda_2 \mathbb{E}[g_2(X)]$$

for any $\lambda_1, \lambda_2 \in \mathbb{R}$, g_1, g_2 integrable

ii) For g_1, g_2 s.t $g_1(x) \leq g_2(x)\ \forall x \in \mathbb{R}$

$$\Rightarrow \quad \mathbb{E}[g_1(X)] \leq \mathbb{E}[g_2(X)] \quad \text{(monotonicity)}$$

iii) If $X \perp\!\!\!\perp Y \Rightarrow \mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$

iv) For $X \geq 0$ (ie, $\mathbb{P}[X<0]=0$)

$$\mathbb{E}[X] = \int_0^\infty (1-F(t))\,dt$$

In particular, if $X \in \mathbb{N}$, then

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}[X \geq n]$$

Mean, variance, moments

- Mean $\mu = \mathbb{E}[X]$

  Variance $\sigma^2 = \mathbb{E}[(X-\mu)^2]$

  $$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- Raw moments - $m_k = \mathbb{E}[X^k]$

  Centered moments - $\sigma_k = \mathbb{E}[(X-\mu)^k]$

- If $X_1 \perp\!\!\!\perp X_2$, $Y = X_1 + X_2$

  then $\sigma_k(Y) = \sigma_k(X_1) + \sigma_k(X_2)$

  (However, for mean, this is always true thanks to linearity!)

Some useful facts about integration

- An important part of analysis/measure theory is to formalize the notion of the integral. For our class, we will just need some important results that come out of this formalism
- Consider a sequence of random variables $X_1, X_2, \ldots$

i) If $X_n(\omega) \geq 0$ a.s. and $X_n(\omega) \leq X_{n+1}(\omega)$ a.s. for all n, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right]$$

(monotone convergence)

ii) If $|X_n(\omega)| \leq Y(\omega)$ a.s. for all n, and $\mathbb{E}[|Y|] < \infty$, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right]$$

(dominated convergence)

iii) If $|X_n(\omega)| \leq c$ a.s for all n, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right]$$

(bounded convergence)

iv) If $X_n \geq Y$ a.s for all n, and $\mathbb{E}[|Y|] < \infty$, then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n]$$

(Fatou's Lemma)

All the above hold for discrete r.v. (and series) as well - see chapter 4.1.3 of Brémaud DP. (Also see Appendix A1 for proofs)

- Interchanging sums and integrals
  - If $X_n \geq 0$ then $\mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty} \mathbb{E}[X_n]$

    If $X(t) \geq 0$ then $\mathbb{E}\left[\int_{-\infty}^{\infty} X(t)\,dt\right] = \int_{-\infty}^{\infty} \mathbb{E}[X(t)]\,dt$

    (Tonelli's Thm)
  - If $\mathbb{E}\left[\sum_{n=1}^{\infty} |X_n|\right] < \infty$, then $\mathbb{E}\left[\sum X_n\right] = \sum \mathbb{E}[X_n]$

    If $\mathbb{E}\left[\int_{-\infty}^{\infty} |X(t)|\,dt\right] < \infty$, then $\mathbb{E}\left[\int X(t)\,dt\right] = \int \mathbb{E}[X(t)]\,dt$

    (Fubini's Thm)