(7) Sperner's Theorem

Def A family $\mathcal{F}$ of subsets of $\{1,\dots,n\}$ is called an antichain if no member of $\mathcal{F}$ is contained in another.

Ex: $\mathcal{F} = \{\text{all subsets of cardinality } k\}$

THM [Sperner] $\forall$ antichain satisfies $|\mathcal{F}| \le \binom{n}{\lfloor n/2 \rfloor}$

Remark: equality achieved for $k = \lfloor n/2 \rfloor$

Proof Consider a random chain

$$C_\sigma := \{\sigma(\{1,\dots,i\}) : i \in \{0,\dots,n\}\}$$

where $\sigma \sim \mathrm{Unif}(S_n)$ is a random permutation

$\forall$ antichain $\mathcal{F}$ can contain at most 1 set from any given chain $\Rightarrow$

$$|\mathcal{F} \cap C_\sigma| \le 1.$$

On the other hand,

$$X := |\mathcal{F} \cap C_\sigma| = \sum_{F \in \mathcal{F}} \mathbb{1}_{\{F \in C_\sigma\}}$$

$$\Rightarrow \quad 1 \ge \mathbb{E}X = \sum_{F \in \mathcal{F}} \mathbb{P}\{F \in C_\sigma\} = \sum_{F \in \mathcal{F}} \frac{1}{\binom{n}{|F|}} \ge |\mathcal{F}| \Big/ \binom{n}{\lfloor n/2 \rfloor}. \qquad \square$$

C_σ contains exactly 1 set of size $|F|$, and this set is uniformly distributed among all $|F|$-element subsets of $\{1,\dots,n\}$

$\binom{n}{\lfloor n/2 \rfloor}$: the middle binomial coefficient is the largest.

(8) Littlewood–Offord Inequality

- If $\pm$ signs are chosen at random, independently,

$$P\{\underbrace{\pm1\pm1\cdots\pm1}_{n\text{ (even)}}=0\}=P\{\#\text{pluses}=\tfrac{n}{2}\}=2^{-n}\binom{n}{n/2}\sim\frac{1}{\sqrt{n}}.$$

- If 1's are replaced with $\forall$ nonzero weights, even better:

THM [Littlewood–Offord; Erdös]

Let $a_1,\dots,a_n$ be nonzero numbers,
$\varepsilon_1,\dots,\varepsilon_n$ be independent Rademacher random var's.

Then $$P\Big\{\sum_1^n \varepsilon_i a_i=0\Big\}\le 2^{-n}\binom{n}{\lfloor n/2\rfloor}\sim\frac{1}{\sqrt{n}}$$

Equivalently, there are $\le\binom{n}{\lfloor n/2\rfloor}$ choices of signs (ε_i) such that $\sum_1^n\varepsilon_i a_i=0$ $(*)$

Proof WLOG $a_i>0$.

- Consider all choices of signs (ε_i) such that $\sum_1^n\varepsilon_i a_i=0$.

  Each such choice is uniquely determined by the set
  $$F:=\{i:\varepsilon_i=+1\}\subset\{1,\dots,n\}.$$

- These sets form an antichain

  (Indeed, if $F\subset F'$, then one can change some "$-$" signs to "$+$" signs and still have the identity

  F: $+++-----$
  F': $+++++---$

  But this is impossible: since all $a_i>0$, the sum must get bigger.)

- Hence, by Sperner's theorem, $\exists$ at most $\binom{n}{\lfloor n/2\rfloor}$ such sets F.

  Each F corresponds one-to-one to a choice of (ε_i) satisfying $\sum_1^n\varepsilon_i a_i=0$

  $\Rightarrow$ $(*)$ holds $\square$.

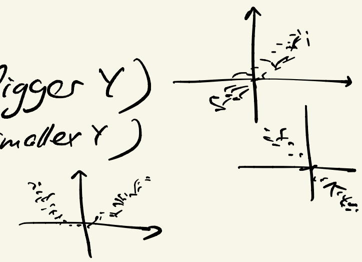
COVARIANCE

Def For random variables X, Y,

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

(proof: $LHS = \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] = \mathbb{E}\left[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y\right]$
$= \mathbb{E}[XY] - \mu_X\mu_Y - \cancel{\mu_Y\mu_X} + \cancel{\mu_X\mu_Y}$ $\square$)

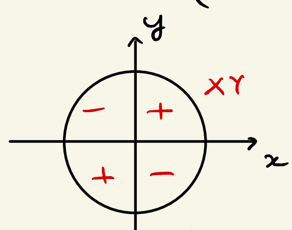
<u>Remarks</u> : 1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$



2. $\mathrm{Cov}(X,Y) > 0$: positively correlated (bigger X tends to imply bigger Y)
   < 0 : negatively correlated (- - — — - — smaller Y)
   $= 0$: uncorrelated

3. X, Y are independent $\Rightarrow$ uncorrelated ($\mathbb{E}[XY] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$)

4. But not vice versa :
   $(X,Y) \sim \mathrm{Unif}(\text{unit disc})$



   $\mathbb{E}X = \mathbb{E}Y = \mathbb{E}[XY] = 0$
   but X, Y are dependent : $X > \frac{1}{\sqrt{2}}$ and $Y > \frac{1}{\sqrt{2}}$ are disjoint events

5. $\boxed{\mathrm{Cov}(X,Y) \le \sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}$ (apply Cauchy-Schwarz for $X-\mathbb{E}X$, $Y-\mathbb{E}Y$)

6. **Def** Pearson correlation coefficient

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\,\sqrt{\mathrm{Var}(Y)}}$$

Unit-free

$\rho_{X,Y} \in [-1,1]$; $\rho_{X,Y} = \pm 1 \iff X = Y$ a.s (by equality condition in Cauchy–Schwarz)

7. $\forall$ r.variables $X_1, \dots, X_n$: $\boxed{\mathrm{Cov}\left(\sum_{i=1}^n X_i\right) = \sum_{i,j=1}^n \mathrm{Cov}(X_i, X_j)}$ ("distributive law")

(proof: WLOG $\mathbb{E}X_i = 0 \Rightarrow LHS = \mathbb{E}\left[\left(\sum_{i,j=1}^n X_i\right)^2\right] = \sum_{i,j=1}^n \mathbb{E}[X_i X_j] = RHS$)

WARNING: Correlation ↙even dependence! does not imply causation

Examples: (a) Midde ages belief: "lice (X) are beneficial for your health (Y)" since they are rarely seen on sick people

"People get sick because lice left them" $(X \Rightarrow Y)$

Reality: vice versa. lice are sensitive to body temperature; fever causes them to leave.

(b) NYC study (1980's):

$\exists$ a strong correlation between # crimes committed (X) and the amount of ice-cream sold by street vendors (Y).

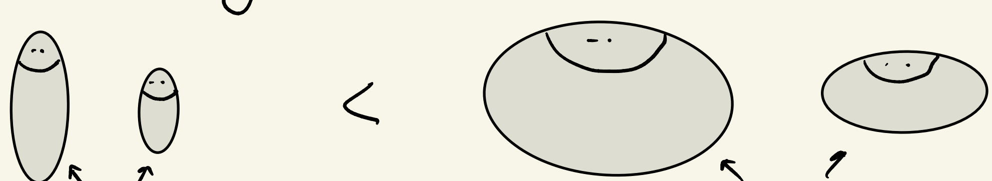
Reality: both crime & ice-cream consumption spike in the summer.

$X \leftarrow Z \rightarrow Y$ Z (time of year) Graphical model. Hidden Markov models.

- BMI Paradox: If you have an average height and average weight, you are overweight.

Let W, H be the weight & height of a randomly chosen person.

- $\boxed{I := \frac{W}{H^3}}$ is the "body proportionality index" (same for $\forall$ homothetic copies)



same I same I

- Assume that $\boxed{I \perp\!\!\!\perp H}$ (taller people are not consistently skinnier/fatter than shorter people)

- If your height $= \mathbb{E}H$, your weight $= \mathbb{E}W$, then your index is

$$I = \frac{\mathbb{E}W}{(\mathbb{E}H)^3} = \frac{\mathbb{E}[IH^3]}{(\mathbb{E}H)^3} = \mathbb{E}[I]\underbrace{\frac{\mathbb{E}[H^3]}{(\mathbb{E}H)^3}}_{\geq 1 \text{ (Jensen)}} > \mathbb{E}[I]. \Rightarrow \text{Overweight}$$

- But the real data indicate that $I < \mathbb{E}[I]$.

So it must be that $I \not\perp\!\!\!\perp H$, specifically I, H must be negatively correlated (taller $\Rightarrow$ skinnier)

- Remark: BMI is defined as $\boxed{\frac{W}{H^2}}$; same paradox occurs.

↖why? H^2 = surface area. fat?