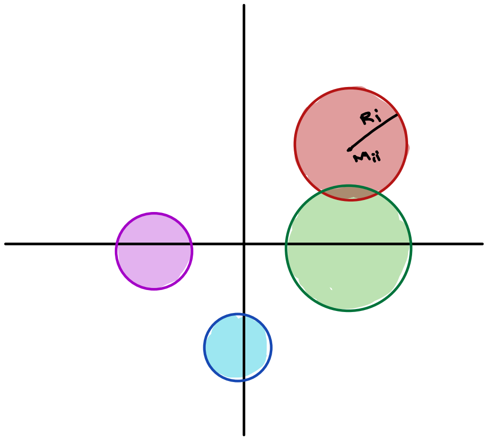
Recap

Gershgorin:



$M \in \mathbb{C}^{n \times n}$

eigenvalue $\lambda \in \bigcup_i \text{Disc}(M_{ii}, R_i)$

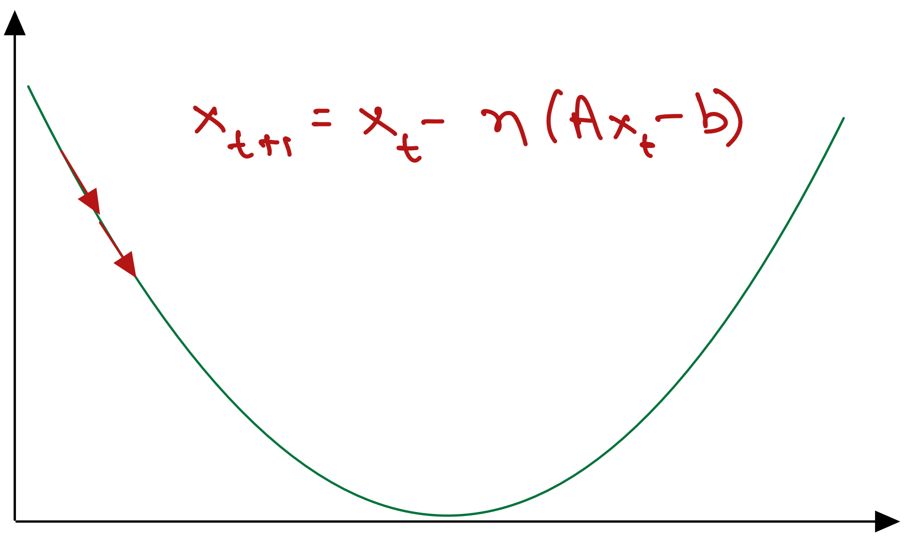
$$R_i = \sum_{j \neq i} |M_{ij}|$$

Linear Systems:

$A_0 x = b_0$

$\downarrow$

$$\underbrace{A_0^T A_0}_{\text{positive definite}} x = A_0^T b_0$$



$$x_{t+1} = x_t - \eta(Ax_t - b)$$

$$(x_t - u) = (I - \eta A)^t (x_0 - u)$$

$$\|(I - \eta A)^t\|_2 = \left(1 - \frac{2}{\kappa + 1}\right)^t$$

$$\left.\kappa = \frac{\lambda_1}{\lambda_n}\right\} \text{condition number}$$

Conjugade gradient method

Steepest descent : $x_0 = 0$ (say)

$$x_1 = x_0 - \eta(Ax_0 - b) = \eta b$$

$$x_2 = x_1 - \eta(Ax_1 - b) = 2\eta b - \eta^2 Ab$$

$$x_3 = x_2 - \eta(Ax_2 - b) = 3\eta b - 3\eta^2 Ab + \eta^3 A^2 b$$

$$\vdots$$

$$x_t \in \underbrace{\mathrm{Span}(\{b, Ab, A^2 b, \ldots, A^{t-1} b\})}_{\text{Krylov space } K_t(A,b)}$$

What if we just found the "best vector" in $K_t(A,b)$

Solve $A_0 x = b_0 \to \underbrace{A_0^T A_0}_{A} x = \underbrace{A_0^T b_0}_{b}$

Say u is unique soln to above

Minimizing in Krylov Space

$$x_t = \underset{x \in K_t(A,b)}{\arg\min} f(x) = \underset{x \in K_t(A,b)}{\arg\min} \|A_0 x - b_0\|^2 = \langle (x-u), A(x-u) \rangle$$

$$= \langle x-u, x-u \rangle_A$$

Let $w_1 \ldots \ldots w_t$ be orthonormal basis for $K_t(A,b)$

under inner product $\langle x, y \rangle_A = \langle x, Ay \rangle$

Let optimal solution $x = \sum_{i=1}^{t} c_i w_i$

Then $c_i = \langle u, w_i \rangle_A = \langle u, Aw_i \rangle = \langle \underline{\underline{Au}}, w_i \rangle$

$= \langle b, w_i \rangle$

Convergence analysis

$$x_t = c_0 \cdot b + c_1 \cdot Ab + \cdots + c_{t-1} \cdot A^{t-1} b$$

$$= \Big(\underbrace{c_0 \cdot I + c_1 \cdot A + c_2 \cdot A^2 \cdots + c_{t-1} \cdot A^{t-1}}_{p(A)}\Big) \cdot \underbrace{b}_{Au}$$

Say $x_0 = 0$

$$u - x_t = (I - p(A) \cdot A) \cdot u = (\underbrace{I - p(A) \cdot A}_{q(A)})(u - x_0)$$

earlier $\quad u - x_t = (I - \eta A)^t (u - x_0)$

Eigenvalues

$$A : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$$

$q(A)$: $q(\lambda_1)$ $\;\cdots\;$ $q(\lambda_n)$

$$\|q(A)\|_2 = \max_{i \in [n]} |q(\lambda_i)|$$

Error analysis

$$(u - x_t) = q(A)\cdot(u - x_0) \Rightarrow \|u - x_t\| \le \|q(A)\|_2 \cdot \|u - x_0\|$$

$$= \max_{i\in[n]} |q(\lambda_i)|$$

Which polynomial $q(A)$ do we choose?

$$q(A) = I - A\cdot p(A)$$

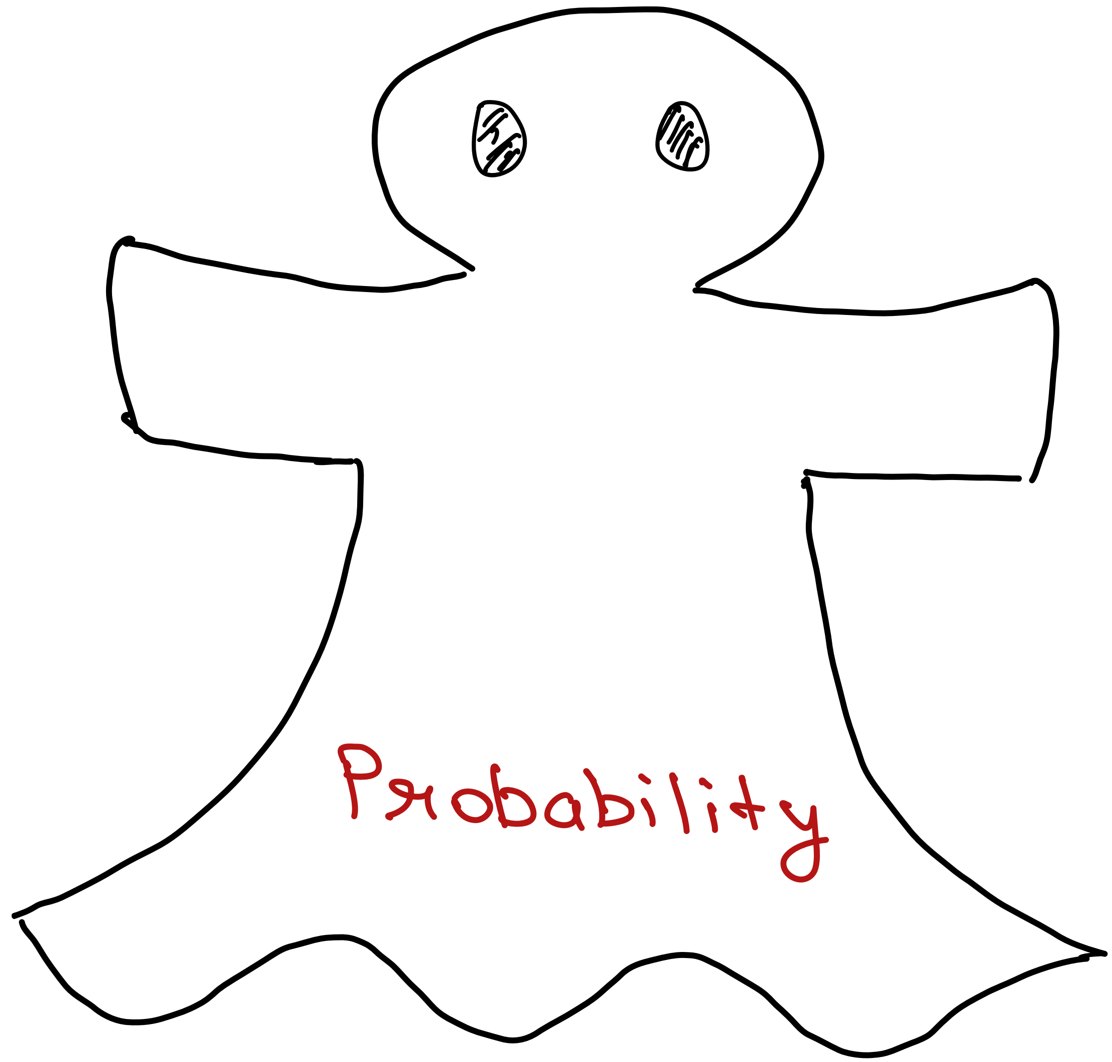
$$q(z) = 1 - z\cdot p(z)$$

- $\exists$ polynomial q, $\deg(q) \le t$, $q(0) = 1$ s.t.

$$q(z) \le 2\cdot\left(1 - \frac{2}{\sqrt{\kappa}+1}\right)^t \quad \forall z \in [\lambda_1, \lambda_n]$$

For $\|q(A)\|_2 \le \epsilon$ choose $t = O(\sqrt{\kappa}\log 1/\epsilon)$

Part II :

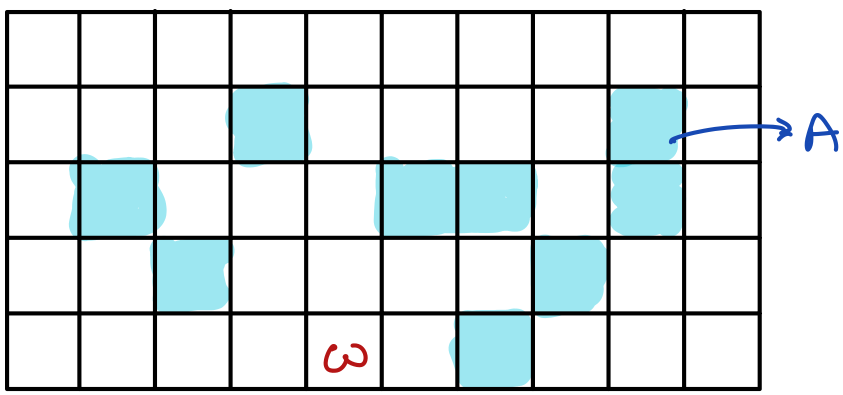


Why?

- Expectations (formally)
- Distributions
- ML / Information Theory

(Finite) Probability Spaces

Specified by a set Ω, and a "measure" $\nu : \Omega \to [0,1]$



$\nu : \Omega \longrightarrow [0,1]$

s.t. $\sum_{\omega \in \Omega} \nu(\omega) = 1$

"probability measure"
"distribution"

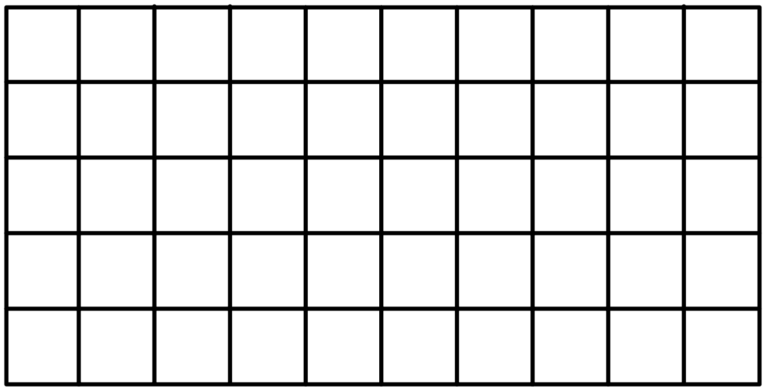
Ω "sample space"
"outcome space"

Event: A subset of outcomes. $A \subseteq \Omega$

$$\mathbb{P}[A] = \sum_{\omega \in A} \nu(\omega)$$

$\{\omega\} \subseteq \Omega$
"atomic" or "elementary"
events

Random Variables (r.v. s)



Function on Ω

A (real-valued) random variable is a function

$$X : \Omega \longrightarrow \mathbb{R}$$

Events $\equiv$ $\underbrace{\{0,1\}\text{-valued r.v.s}}_{\text{indicator r.v.s}}$

$$A \subseteq \Omega \equiv \chi_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Expectations

For probability space (Ω, ν), r.v. $X : \Omega \to \mathbb{R}$

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} \nu(\omega)\, X(\omega) = \sum_{a} \underbrace{\left(\sum_{\omega : X(\omega) = a} \nu(\omega) \right)}_{\mathbb{P}[X = a]} \cdot a$$

Linearity of expectation

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y], \quad \mathbb{E}[c \cdot X] = c \cdot \mathbb{E}[X]$$

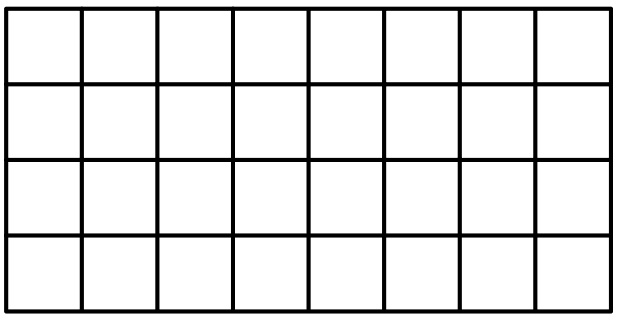
$V = \{X : \Omega \to \mathbb{R}\}$ is a vector space

$\mathbb{E} : V \to \mathbb{R}$ is a linear transformation

Shuffling Cards

Shuffle new deck of n cards to get a "random order"

How many cards in same place as before?



$X : \Omega \longrightarrow \mathbb{R}$ = # cards in same place.

$E[X] = E[X_1 + \dots \quad \dots + X_n]$

Ω = Set of all $n!$ orderings

$\nu(\omega) = \dfrac{1}{n!}$

This is a common choice, but certainly not the only one!

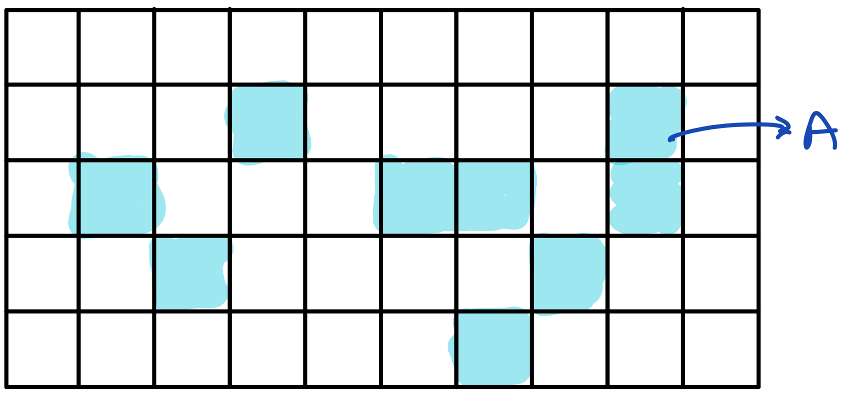
X_i = i^{th} card in same place

$E[X_i] = \dfrac{(n-1)!}{n!} = \dfrac{1}{n}$

$E[X] = 1$

Conditioning

Conditioning on event $A \subseteq \Omega$ $\equiv$ Restricting probability space to A



$$\nu_A(\omega) = \begin{cases} 0 & \text{if } \omega \notin A \\ \dfrac{\nu(\omega)}{\mathbb{P}[A]} & \text{if } \omega \in A \end{cases}$$

$$\mathbb{P}[B \mid A] = \sum_{\omega \in B} \nu_A(\omega)$$

$$= \sum_{\omega \in B} \frac{\nu(\omega)}{\mathbb{P}[A]} \cdot \mathbb{1}_A(\omega)$$

$$= \sum_{\omega \in A \cap B} \frac{\nu(\omega)}{\mathbb{P}(A)} = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}$$

$$\mathbb{E}[X \mid A] = \sum_{\omega} \nu_A(\omega)\, X(\omega)$$

$$= \frac{\mathbb{E}[\mathbb{1}_A(\omega) \cdot X(\omega)]}{\mathbb{P}[A]}$$