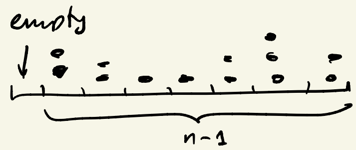
Application:

Occupancy Problem Drop m balls into n bins independently at random. What is the distribution of the fraction of empty bins?

- $\#(\text{empty bins}) =: S_n = X_1 + \dots + X_n$ where $X_i = \begin{cases} 1 & \text{if } \underline{\text{bin } i \text{ is empty}} \\ 0, & \end{cases}$

  $X_i \sim \mathrm{Ber}\left(\left(1 - 1/n\right)^m\right)$ <u>NOT</u> independent, but



- $\mathbb{E} S_n = \sum_{i=1}^n \mathbb{E} X_i = n\, \mathbb{P}\{X_1 = 1\} = n\left(\frac{n-1}{n}\right)^m$

Asymptotic behavior? Let $m = m(n)$, $\frac{m}{n} \to \lambda$ as $n \to \infty$

"oversampling factor"

$$\Rightarrow \quad \frac{\mathbb{E} S_n}{n} = \left(1 - \frac{1}{n}\right)^{\lambda n} \to e^{-\lambda}. \qquad \text{(decays fast!)}$$

THM $\frac{S_n}{n} \to e^{-\lambda}$ in probability

Proof: By WLLN (remark p. 78), it suffices to check that X_i are <u>negatively correlated</u>.

Heuristically, this is true: an empty bin increases other bins to be full.

Formally:

$$\mathrm{Cov}(X_i, X_j) = \underbrace{\mathbb{E}[X_i X_j]}_{= \left(\frac{n-2}{n}\right)^m} - \underbrace{\mathbb{E}[X_i] \cdot \mathbb{E}[X_j]}_{= \left(\frac{n-1}{n}\right)^{2m}}$$

$$\left(\frac{n-2}{n}\right)^m \overset{?}{<} \left(\frac{n-1}{n}\right)^{2m}$$

$$\Updownarrow$$

$$1 - \frac{2}{n} < \left(1 - \frac{1}{n}\right)^2 = 1 - \frac{2}{n} + \frac{1}{n^2}. \qquad \text{True!} \quad \square.$$

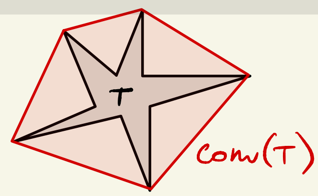
Remark: The WLLN does not apply directly since the distributions of each X_i depend on n ($\Rightarrow$ we have a "triangular array" of $X_i^{(n)}$)

But the <u>proof</u> of WLLN works fine. $\longrightarrow$ **HW**.

Example: Approximate Caratheodory Theorem [see my HDP book]

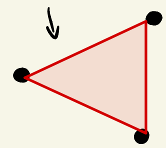
Def The convex hull of a set $T \subset \mathbb{R}^d$, denoted $\mathrm{conv}(T)$, is the smallest convex set that contains T. Equivalently,

$$x \in \mathrm{conv}(T) \iff x = \underbrace{\sum_{i=1}^{m} \lambda_i z_i}_{\text{convex combination}} \text{ for some } \lambda_i \geq 0, \ \sum_{i=1}^{m} \lambda_i = 1, \ z_i \in T$$



Thm (Caratheodory) $m = d+1$ suffices.

and is optimal in general



But if we allow to approximate x, we get a dramatic improvement:

Thm (Approximate Caratheodory) Let $T \subset$ unit ball of $\mathbb{R}^d$.

Then $\forall x \in \mathrm{conv}(T)$, $\forall k \in \mathbb{N}$ $\exists z_1, \dots, z_n \in T$:

$$\left\| x - \frac{1}{n} \sum_{i=1}^{n} z_i \right\|_2 \leq \frac{1}{\sqrt{n}}.$$

- Dimension-free; does not depend on geometry of T !

<u>Proof</u> ("Empirical method", B. Maurey)

- Fix $x \in \operatorname{conv}(T)$, express it as a convex combination

$$x = \sum_{i=1}^{m} \lambda_i z_i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^{m} \lambda_i = 1, \quad z_i \in T.$$

potentially large

- Interpret λ_i as probabilities, the sum as expectation

  let r.vector Z take value z_i with prob. λ_i

$$\Rightarrow \quad x = \mathbb{E} Z$$

- Consider indep. copies $Z_1, Z_2, \ldots$ of Z.

  SLLN: $\frac{1}{n} \sum_{i=1}^{n} Z_i \to \mathbb{E} Z = x$ a.s.

- Error:

$$\mathbb{E} \left\| x - \frac{1}{n} \sum_{i=1}^{n} Z_i \right\|_2^2 = \mathbb{E} \left\| \frac{1}{n} \sum_{i=1}^{n} (Z_i - \underbrace{x}_{= \mathbb{E} Z_i}) \right\|_2^2$$

Analog of $\operatorname{Var}(\text{sum}) = \text{sum}(\text{Variances})$ in $\mathbb{R}^d$ (DIY)

$$= \frac{1}{n^2} \sum_{i=1} \underbrace{\mathbb{E} \| Z_i - \mathbb{E} Z_i \|_2^2}_{= \mathbb{E} \| Z - \mathbb{E} Z \|_2^2}$$

analog of $\operatorname{Var}(Z) = \mathbb{E} Z^2 - (\mathbb{E} Z)^2$ (DIY)

$$= \frac{1}{n} \mathbb{E} \| Z - \mathbb{E} Z \|_2^2 = \frac{1}{n} \Big(\underbrace{\mathbb{E} \| Z \|^2}_{\leq 1 \text{ (asm.)}} - \underbrace{\| \mathbb{E} Z \|_2^2}_{\geq 0} \Big) \leq \frac{1}{n}.$$

- $\Rightarrow \exists$ realization of the r.v's $Z_1, \ldots, Z_n$ s.t.

$$\left\| x - \frac{1}{n} \sum_{i=1}^{n} Z_i \right\|_2^2 \leq \frac{1}{n}$$

Since $Z_i \in T$, QED $\lrcorner$

Application: polytopes with few vertices have exponentially small volume:

THM [Carl–Pajor '88]

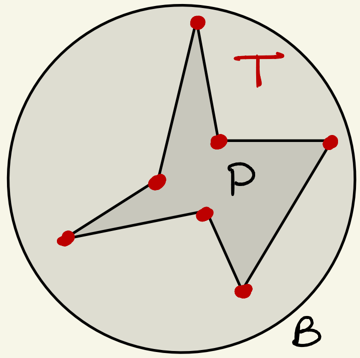
Let $P \subset B$ polytope with m vertices

unit Euclidean ball

Then $\dfrac{\mathrm{Vol}(P)}{\mathrm{Vol}(B)} \le \left(3\sqrt{\dfrac{\log m}{d}}\right)^d.$

$\wedge$

2^{-d} if $m < \exp(cd)$



Proof

- $P \subset \mathrm{conv}(T)$ where $T = \{\text{vertices of } P\} \subset B$
- A.C.T $\Rightarrow$

  $\forall x \in P \subset \mathrm{conv}(T)$ is within distance $\frac{1}{\sqrt{n}}$ from some point in the set

  $$\mathcal{N} := \left\{ \frac{1}{n} \sum_{i=1}^{n} x_i \;:\; x_i \in T \right\}$$

  $\Rightarrow$ $\forall x \in P$ is covered by a ball of radius $\frac{1}{\sqrt{n}}$ and center $\in \mathcal{N}$. The number of such balls is

  $$|\mathcal{N}| \le m^n$$

  (#ways to choose n elements x_i from the set T of m elements, with repetition)

  $\Rightarrow$ P can be covered by m^n copies of $\frac{1}{\sqrt{n}}B$.

- Comparing volumes $\Rightarrow$ $\mathrm{vol}(P) \le m^n \,\mathrm{Vol}\left(\frac{1}{\sqrt{n}}B\right) = m^n \left(\frac{1}{\sqrt{n}}\right)^d \mathrm{Vol}(B)$

  $$\frac{\mathrm{Vol}(P)}{\mathrm{Vol}(B)} \le \frac{m^n}{n^{d/2}} \le \left(3\sqrt{\frac{\log m}{d}}\right)^d.$$

  (optimize n: choose $n = \dfrac{d}{2\log m}$)

$\square$.

Remarks

① A slightly more accurate count yields the bound
$$\left(C\sqrt{\frac{\log(m/d)}{d}}\right)^d.$$

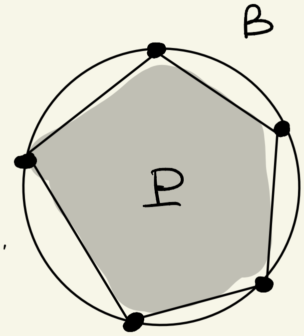
② It is optimal, attained by a random polytope

$$P = \mathrm{conv}\{x_1, \dots, x_m\}$$

where x_i = indep. random pts on the unit sphere.

[Dafnis-Giannopoulos-Tsolomitis '2003, 2009]

— see next page



hyperbolic intuition

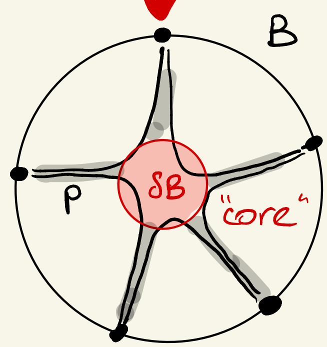
③ "Hyperbolic intuition"

for $\delta := 3\sqrt{\frac{\log m}{d}}$, Thm says:

tiny ball!

$$\mathrm{Vol}(P) \le \delta^d\, \mathrm{Vol}(B) = \mathrm{Vol}(\delta B).$$

Thus the picture "actually" looks like this, even though it violates convexity.



V. Milman's "hyperbolic intuition"

④ A "blind spider" experiment:

- A random walk will stay in the "core" δB, will not find the "tentacles".

- The spider can't tell the polytope from the ball δB.

- Bad for algorithms (MCMC)