1. Concentration inequalities

Thursday, September 9, 2021 2:47 PM

Key idea: Often, behavior of a r.v. is not too far off from its expected behavior.

Thm. 2.1 (Markov's inequality)

Let x be a nonnegative r.v. Then, for $a > 0$,

$$\text{Prob}(x \geq a) \leq \frac{\mathbb{E}x}{a}.$$

proof. Let x be continuous with prob. density p. (similar for discrete)

$$\mathbb{E}x = \int_0^\infty x\,p(x)\,dx = \int_0^a x\,(p(x))\,dx + \int_a^\infty x\,p(x)\,dx$$

$$\geq \int_a^\infty x\,p(x)\,dx \geq a\int_a^\infty p(x)\,dx = a\,\text{Prob}(x \geq a)$$

since $x \in [a, \infty)$

$$\Rightarrow \text{Prob}(x \geq a) \leq \frac{\mathbb{E}x}{a} \qquad \boxed{\diagup}$$

Corr. $\text{Prob}(x \geq b\,\mathbb{E}x) \leq \frac{1}{b}$.

Thm. 2.3 (Chebyshev Inequality)

Let x be a r.v. with bounded variance. Then for $c > 0$,

$$\text{Prob}(|x - \mathbb{E}x| \geq c) \leq \frac{\text{Var}(x)}{c^2}.$$

proof $\text{Prob}(|x - \mathbb{E}x| \geq c) = \text{Prob}(|x - \mathbb{E}x|^2 \geq c^2)$

Let $y = |x - \mathbb{E}x|^2$. Then y is a nonnegative r.v. and $\mathbb{E}y = \text{Var}(x)$.

By Markov, $\text{Prob}(|x - \mathbb{E}x| \geq c) = \text{Prob}(y \geq c^2) \leq \frac{\mathbb{E}y}{c^2} = \frac{\text{Var}(x)}{c^2}$ $\boxed{\diagup}$

Thm 2.4 (Law of Large Numbers)

Let $x_1, \dots, x_n$ be i.i.d. samples of a r.v. x. Then

$$\text{Prob}\left(\left|\frac{x_1 + \dots + x_n}{n} - \mathbb{E}x\right| \geq \varepsilon\right) \leq \frac{\text{Var}(x)}{n\varepsilon^2}$$

Proof. By Chebyshev,

$$\text{Prob}\left(\left|\frac{x_1+\cdots+x_n}{n} - \mathbb{E}x\right| \geq \varepsilon\right) \leq \frac{\text{Var}\left(\frac{x_1+\cdots+x_n}{n}\right)}{\varepsilon^2}$$

$$= \frac{1}{n^2\varepsilon^2}\text{Var}(x_1+\cdots+x_n)$$

$$= \frac{1}{n^{2}\varepsilon^2}\left[\text{Var}(x_1)+\cdots+\text{Var}(x_n)\right] = \frac{\text{Var}(x)}{n\varepsilon^2}. \quad \square$$

Consider random Gaussian pts in $\mathbb{R}^d$

i.e. Let $\vec{y} = [y_1, \ldots, y_d] \in \mathbb{R}^d$, y_i's and z_i's are i.i.d.

$\vec{z} = [z_1, \ldots, z_d] \in \mathbb{R}^d$ with pdf $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

(i.e. $y_i \sim \mathcal{N}(0,1)$)

What can we say about $|y-z|$?

$|y-z|^2 = \sum_{i=1}^{d}(y_i - z_i)^2$. Let $x_i = (y_i - z_i)^2$ a r.v. w/ bounded variance.

By LLN, $\text{Prob}\left(\left|\frac{x_1+\cdots+x_d}{d} - \mathbb{E}x\right| \geq \varepsilon\right) \leq \frac{\text{Var}(x)}{d\varepsilon^2}$.

Thm. 2.5 (Master Tail Bounds Thm)

Let $x = x_1 + \cdots + x_n$, where $x_1, \ldots, x_n$ are mutually ind. r.v. with 0 mean and variance at most σ^2. Let $0 \leq a \leq \sqrt{2}\, n\sigma^2$.

Assume that $|\mathbb{E}(x_i^s)| \leq \sigma^2 s!$ for $s = 3, 4, \ldots, \lfloor \frac{a^2}{4n\sigma^2} \rfloor$.

Then $\text{Prob}(|x| \geq a) \leq 3e^{-a^2/12n\sigma^2}$.

proof sketch: Apply Markov to x^r, r large and even. (so $x^r \geq 0$)

$$\text{Prob}(|x| \geq a) = \text{Prob}(x^r \geq a^r) \leq \frac{\mathbb{E}x^r}{a^r}.$$

Compute $\mathbb{E}x^r$ by expanding out $x = x_1 + \cdots + x_n$ and using properties of the expectation and the technical assumptions.

Will find $\mathbb{E}x^r$ not too large. $\square$

properties of the expectation and the technical assumptions.

Will find $\mathbb{E}x^r$ not too large. $\boxed{\diagup}$

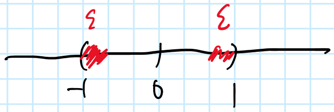
More useful tail bounds:

	Cond	Tail Bound
Chernoff	$x = x_1 + \dots + x_n$, $x_i \in \{0, 1\}$ i.i.d. Bernoulli	$\text{Prob}(\lvert x - \mathbb{E}x\rvert \geq \varepsilon\, \mathbb{E}x) \leq 3e^{-c\varepsilon^2 \mathbb{E}x}$
Higher Moments	r pos even int.	$\text{Prob}(\lvert x\rvert \geq a) \leq \dfrac{\mathbb{E}\,x^r}{a^r}$ (Markov)
Gaussian Annulus	$x = \sqrt{x_1^2 + \dots + x_n^2}$, $x_i \sim \mathcal{N}(0,1)$; $\beta \leq \sqrt{n}$	$\text{Prob}(\lvert x - \sqrt{n}\rvert \geq \beta) \leq 3e^{-c\beta^2}$.

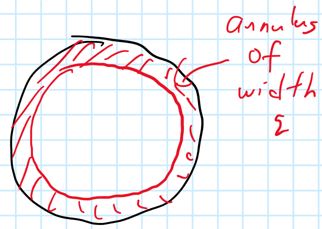
Many variants

Most volume of high-dim objects is near the surface.

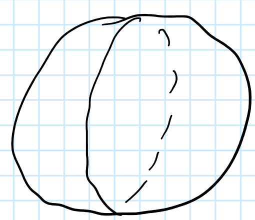
Ex.



Vol frac $\frac{2\varepsilon}{2} = \varepsilon$



Vol frac: $\approx \frac{2\pi\varepsilon}{\pi} = 2\varepsilon$



Vol frac. $\approx \frac{4\pi\varepsilon}{\frac{4}{3}\pi} = 3\varepsilon$

More rigorously, consider any object $A \subset \mathbb{R}^d$.

Shrink A by ε to produce $(1-\varepsilon)A = \{(1-\varepsilon)x \mid x \in A\}$.

Then $\text{vol}((1-\varepsilon)A) = (1-\varepsilon)^d\, \text{vol}(A)$.

proof sketch: partition A into infinitesimal hypercubes. Then $(1-\varepsilon)A$ is the union of the set of cubes obtained by shrinking the cubes by a factor of $(1-\varepsilon)$.

Shrinking side-lengths by $(1-\varepsilon)$ implies shrinking vol by $(1-\varepsilon)^d$. $\boxed{\diagup}$

Shrinking side-lengths by $(1-\varepsilon)$ implies shrinking vol by $(1-\varepsilon)^d$. $\square$

Note that $1-x \le e^{-x}$, so for any $A \subset \mathbb{R}^d$,

$$\frac{\mathrm{vol}\left((1-\varepsilon)A\right)}{\mathrm{vol}(A)} = (1-\varepsilon)^d \le e^{-\varepsilon d} \longrightarrow 0 \quad \text{as} \quad d \to \infty.$$

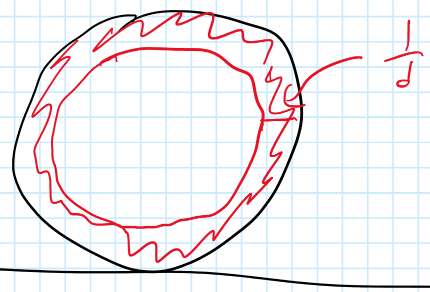
Thus, most vol does not belong to $(1-\varepsilon)A$.

Going back to the unit ball $B_d \subseteq \mathbb{R}^d$.

Then $\dfrac{\mathrm{vol}\left((1-\varepsilon)B_d\right)}{\mathrm{vol}(B_d)} \le e^{-\varepsilon d}$, so $\mathrm{vol}\left(B_d \setminus (1-\varepsilon)B_d\right) \ge \left(1-e^{-\varepsilon d}\right)\mathrm{vol}(B_d)$.

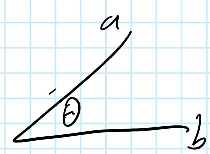
Let $\varepsilon = \frac{1}{d}$. $\mathrm{Vol}\left(B_d \setminus \left(1-\frac{1}{d}\right)B_d\right) \ge \left(1-e^{-1}\right)\mathrm{vol}(B_d) \approx 0.632\, \mathrm{vol}(B_d)$.

$\Rightarrow$ most vol is contained in an annulus of width $\frac{1}{d}$ near the boundary.



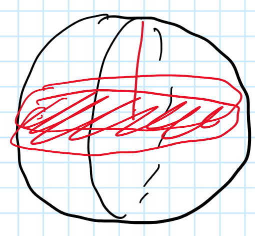
Most points in a unit ball are nearly orthogonal

Recall dot product $\vec{a} \cdot \vec{b} = \sum_{i=1}^{d} a_i b_i = \|a\|\|b\| \cos\theta$



So $a \cdot b$ is small $\approx$ $a \perp b$ are nearly orthogonal.

WLOG, fix $\vec{e}_1$, the first coordinate vector as "North".



Then $\vec{e}_1 \cdot \vec{x} = x_1$.