- INEQUALITIES -

<u>Chebyshev's ineq</u> : if X is a nonneg r.v. then

$$\mathbb{P}(X \geq t) \leq \frac{1}{t}\,\mathbb{E}X, \qquad t > 0.$$

<u>Proof</u>

$$X \geq X\,1_{\{X \geq t\}} \geq t\,1_{\{X \geq t\}}$$

$$\mathbb{E}X \geq \quad \cdots \quad \geq t\,\mathbb{E}1_{\{X\geq t\}} = t\,\mathbb{P}(X \geq t). \square$$

<u>Variants</u> :

- $\mathbb{P}(X \geq t) \underset{p>0}{=} \mathbb{P}(X^p \geq t^p) \leq \frac{1}{t^p}\,\mathbb{E}X^p$

  (p^{th} moment Markov)

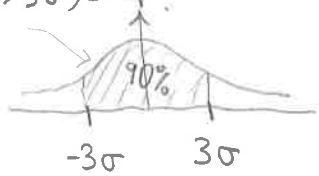
- $\mathbb{P}(X \geq t) \underset{\lambda>0}{=} \mathbb{P}(e^{\lambda X} \geq e^{\lambda t}) \leq \frac{1}{e^{\lambda t}}\,\mathbb{E}e^{\lambda X}$

  (for any $t \in \mathbb{R}$, X r.v.)

  (exponential Cheb)

- $\mathbb{P}(|X - \mathbb{E}X| > t\sqrt{\operatorname{Var}X}) \leq \frac{1}{t^2 \cdot \operatorname{Var}X}\,\mathbb{E}|X - \mathbb{E}X|^2 = \frac{1}{t^2},$

e.g. $t = 3 \rightsquigarrow$ $\mathbb{P}(|X-\mathbb{E}X| > 3\sigma) \leq \frac{1}{9}$.



$\sigma = \sqrt{\operatorname{Var}X}$ $\qquad$ (3σ - rule)

<u>Hölder's ineq</u>

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p}\,(\mathbb{E}|Y|^q)^{1/q},$$

$p, q \geq 1$, $\frac{1}{p} + \frac{1}{q} = 1$

(when $p = 1$, $q = \infty$)

In part. $p = q = 2 \rightsquigarrow$ <u>Cauchy - Schwarz ineq</u>

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}|X|^2}\,\sqrt{\mathbb{E}|Y|^2}.$$

Proof of Hölder's ineq. $\forall x, y \geq 0 \qquad xy \leq \frac{x^p}{p} + \frac{y^q}{q}$

explanation: $\log\left(\frac{1}{p}x^p + \frac{1}{q}y^q\right) \underset{\substack{\uparrow \\ \log \text{ is concave} \\ \frac{1}{p}+\frac{1}{q}=1}}{\geq} \frac{1}{p}\log x^p + \frac{1}{q}\log y^q$

$= \log(xy)$.

Plug in $x = \frac{|X|}{(\mathbb{E}|X|^p)^{1/p}}$, $y = \frac{|Y|}{(\mathbb{E}|Y|^q)^{1/q}}$, take the expectation. $\square$

Recall: p^{th} moment $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$, $p > 0$

(measure how "large" X is)

Thm $0 < p < q \Rightarrow \|X\|_p \leq \|X\|_q$

Proof $\mathbb{E}|X|^p = \mathbb{E}|X|^p \cdot 1 \underset{\substack{\text{Hölder} \\ \frac{1}{r}+\frac{1}{s}=1,\ r=\frac{q}{p}>1}}{\leq} \left(\mathbb{E}(|X|^p)^r\right)^{1/r} (\mathbb{E}1^s)^{1/s}$. $\square$

Minkowski's ineq $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$, $p \geq 1$

Proof We have a variational formula

$$\|X\|_p = \sup\{\mathbb{E}XY, \quad \mathbb{E}|Y|^q \leq 1\}, \quad \frac{1}{p}+\frac{1}{q}=1$$

Explanation: $\mathbb{E}XY \leq (\mathbb{E}|X|^p)^{1/p}(\mathbb{E}|Y|^q)^{1/q} \leq \|X\|_p$,

equality attained for $Y = \operatorname{sgn}(X)\,|X|^{p-1}\,\frac{1}{\|X\|_p^{p/q}}$.

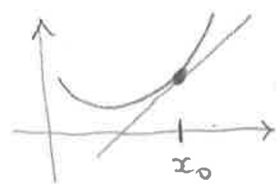
$$\|X+Y\|_p = \sup\Big\{ \underbrace{\mathbb{E}(X+Y)Z}_{\mathbb{E}XZ + \mathbb{E}YZ}, \quad \mathbb{E}|Z|^q \le 1 \Big\}$$

$$\le \sup \mathbb{E}XZ + \sup \mathbb{E}YZ = \|X\|_p + \|Y\|_p . \quad \square$$

<u>Jensen's ineq.</u> If $f:\mathbb{R}\to\mathbb{R}$ is convex, then

$$\mathbb{E}f(X) \ge f(\mathbb{E}X).$$

<u>Proof</u>



$$f(x) \ge f(x_0) + f'(x_0)\cdot(x-x_0)$$

$x = X$, $x_0 = \mathbb{E}X$, take $\mathbb{E}$.

$$\mathbb{E}f(X) \ge f(\mathbb{E}X) + f'(\mathbb{E}X)\,\mathbb{E}(X-\mathbb{E}X) = f(\mathbb{E}X). \ \square$$

<u>E.g.</u> $0<p<q$, $r = \frac{q}{p} > 1$, $x \overset{f}{\mapsto} |x|^r$ convex

$$\mathbb{E}|X|^q = \mathbb{E}f(|X|^p) \ge f(\mathbb{E}|X|^p) = (\mathbb{E}|X|^p)^{q/p} \rightsquigarrow \|X\|_q \ge \|X\|_p .$$

We define $L_p = L_p(\Omega, \mathcal{F}, \mathbb{P}) = \{ X \text{ r.v. s.t. } \mathbb{E}|X|^p < \infty \}$

Minkowski's ineq $\rightsquigarrow$ L_p is a linear space,

$\|\cdot\|_p$ is a norm on L_p meaning

$\|\lambda X\|_p = |\lambda| \cdot \|X\|_p$, $\lambda \in \mathbb{R}$ (homogeneity)

$\|X+Y\|_p \le \|X\|_p + \|Y\|_p$ (triangle ineq.)

E.g. (Bernstein's ineq.) Let $\varepsilon_1, \varepsilon_2, \dots$ be indep. random signs, $a_1, \dots, a_n \in \mathbb{R}$. Then for $t > 0$

$$\mathbb{P}\left(\left| \sum_{i=1}^n a_i \varepsilon_i \right| > t \right) \leq 2 e^{-\frac{t^2}{2\sigma^2}}, \qquad \sigma^2 = \sum a_i^2,$$

Let $S = \sum a_i \varepsilon_i$. We have $\mathbb{P}(|S| > t) = \mathbb{P}(\{S > t\} \cup \{S < -t\})$

$$= \mathbb{P}(S > t) + \mathbb{P}(S < -t) = \mathbb{P}(S > t) + \mathbb{P}(\underbrace{-S}_{\text{the same dist as } S} > t)$$

$$= 2\mathbb{P}(S > t).$$

Now the exp. Chebyshev's ineq. yields

$$\mathbb{P}(S > t) = \mathbb{P}(e^{\lambda S} > e^{\lambda t}) \leq e^{-\lambda t} \mathbb{E} e^{\lambda S},$$

$$\mathbb{E} e^{\lambda S} = \mathbb{E} \prod_{i=1}^n e^{\lambda a_i \varepsilon_i} = \prod_{i=1}^n \mathbb{E} e^{\lambda a_i \varepsilon_i} = \prod_{i=1}^n \left(\frac{e^{\lambda a_i} + e^{-\lambda a_i}}{2} \right)$$

$$\underset{\frac{e^x + e^{-x}}{2} \leq e^{+x^2/2}}{\leq} \prod e^{\lambda^2 a_i^2 / 2} = e^{\frac{\lambda^2}{2} \sum a_i^2} = e^{\frac{\lambda^2}{2} \sigma^2}$$

$$\mathbb{P}(S > t) \leq e^{\frac{\lambda^2}{2}\sigma^2 - \lambda t} \qquad \forall \lambda > 0$$

Choose λ s.t. RHS as small as possible $\rightsquigarrow \lambda = t/\sigma^2$

which gives

$$\mathbb{P}(S > t) \leq e^{-\frac{t^2}{2\sigma^2}}.$$

E.g. $a_i = 1/\sqrt{n}$, $\sigma^2 = 1 = \mathrm{Var}(\sum a_i \varepsilon_i)$, $\mathbb{P}\left(\left| \frac{\sum \varepsilon_i}{\sqrt{n}} \right| > t \right) \leq 2e^{-t^2/2}$

"Gaussian decay as $t \to \infty$"

E.g. Expectation via tail. Let $X \geq 0$

$$\mathbb{E}X = \mathbb{E}\int_0^X dt = \mathbb{E}\int_0^\infty 1_{\{t<X\}}\, dt = \int_0^\infty (\mathbb{E} 1_{\{t<X\}})\, dt$$
$$= \int_0^\infty \mathbb{P}(X>t)\, dt.$$

E.g. If $X \geq 0$ and $\mathbb{E}X < \infty$, then $t\mathbb{P}(X>t) \xrightarrow[t\to\infty]{} 0$.

Take $t_n \nearrow \infty$, $X_n = t_n \mathbb{1}_{\{X>t_n\}}$. We have

$$\mathbb{E}X_n = t_n \mathbb{P}(X>t_n)$$

so WTS $\mathbb{E}X_n \xrightarrow[n\to\infty]{} 0$. Since $X_n \xrightarrow[n\to\infty]{} 0$, X_n are dominated, $X_n \leq X\mathbb{1}_{\{X>t_n\}} \leq X \leftarrow$ integrable, we get by Lebesgue's dominated convergence thm

$$\lim_{n\to\infty} \mathbb{E}X_n = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) = \mathbb{E}0 = 0.$$

E.g. Weierstrass thm: For every $f : [0,1] \to \mathbb{R}$ cts, $\varepsilon > 0$, there is a polynomial P s.t. $\sup_{x\in[0,1]} |f(x) - P(x)| \leq \varepsilon$.

Proof Fix $x \in [0,1]$, $n \geq 1$, let $S_n^x \sim \mathrm{Bin}(n,x)$,

$$B_n(x) = \mathbb{E} f\left(\frac{S_n^x}{n}\right) = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right).$$

(Bernstein poly)

Want $\quad B_n \approx f$. We have

$$f(x) - B_n(x) = \mathbb{E}\, f(x) - f\left(\frac{S_n^x}{n}\right)$$

so $\quad |f(x) - B_n(x)| \leq \mathbb{E}\left| f(x) - f\left(\frac{S_n^x}{n}\right)\right|$

$$= \mathbb{E}\left| f(x) - f\left(\frac{S_n^x}{n}\right)\right| \mathbb{1}\left\{ \left|x - \frac{S_n^x}{n}\right| < n^{-1/4}\right\}$$

$$+ \mathbb{E}\underbrace{\left| f(x) - f\left(\frac{S_n^x}{n}\right)\right|}_{\leq 2M} \mathbb{1}\left\{ \left|x - \frac{S_n^x}{n}\right| \geq n^{-1/4}\right\}$$

Since f is cts,
- f is bdd $\quad |f(x)| \leq M \quad \forall x \in [0,1]$
- f is uniformly cts on $[0,1]$.

$$\forall \varepsilon\ \exists \delta\ \ \forall x, y \in [0,1] \quad |x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$$

Fix $\varepsilon > 0$ and choose δ. For $n > n_0$ s.t. $n_0^{-1/4} < \delta$ we get

$$\mathbb{E}\underbrace{\left| f(x) - f\left(\frac{S_n^x}{n}\right)\right|}_{\leq \varepsilon} \underbrace{\mathbb{1}\left\{ \left|x - \frac{S_n^x}{n}\right| < n^{-1/4}\right\}}_{\leq 1} \leq \varepsilon.$$

By Chebyshev's ineq.,

$$\mathbb{P}\left(\left|x - \frac{S_n^x}{n}\right| \geq n^{-1/4}\right) \leq \frac{1}{(n^{-1/4})^2}\, \mathbb{E}\left| \underset{\mathbb{E}\frac{S_n^x}{n}}{\underset{\parallel}{x}} - \frac{S_n^x}{n}\right|^2$$

$$= \sqrt{n} \cdot \mathrm{Var}\left(\frac{S_n^x}{n}\right)$$

$$= \sqrt{n} \cdot \frac{1}{n^2} \cdot n\, x(1-x) \leq \frac{1}{\sqrt{n}},$$

so, altogether,

$$|f(x) - B_n(x)| \leq 2M \cdot \frac{1}{\sqrt{n}} + \varepsilon \leq 2\varepsilon \quad \text{for } n > n_0'. \ \square$$

<u>E.g.</u> First moment method: suppose we have a nonneg integer-valued r.v. X and want to show that $X=0$ with high probability (w.h.p). We have

$$\mathbb{P}(X>0) = \mathbb{P}(X \geq 1) \overset{\text{Cheb.}}{\leq} \mathbb{E}X.$$

For example, throw m balls uniformly and independently at random into n bins. Show that if $m > (1+\varepsilon) n \log n$, w.h.p. there are no empty bins.

$$X = \text{no. of empty bins} = \sum_{i=1}^{n} X_i, \quad X_i = \begin{cases} 1 & i^{\text{th}} \text{ bin empty} \\ 0 & \text{o/w} \end{cases}$$

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

$$\underset{1+x \leq e^x}{\leq} n \cdot e^{-\frac{m}{n}} < n \cdot e^{-(1+\varepsilon)\log n}$$

$$= n^{-\varepsilon}$$

so $\mathbb{P}(X>0) \leq \mathbb{E}X \leq \frac{1}{n^{\varepsilon}}$, or

$$\mathbb{P}(X=0) = \mathbb{P}(\text{no empty bins}) \geq 1 - \frac{1}{n^{\varepsilon}}.$$

<u>E.g.</u> Second moment method: suppose we want to show that $X>0$ w.h.p. We have (cf. Q5 HW7)

$$\mathbb{E}X \overset{X\ge 0}{=} \mathbb{E}X\mathbb{1}_{\{X>0\}} \overset{\text{C-S}}{\le} \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}\mathbb{1}^2_{\{X>0\}}}$$

$$= \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}\mathbb{1}_{\{X>0\}}} = \sqrt{\mathbb{E}X^2}\cdot\sqrt{\mathbb{P}(X>0)}$$

hence

$$\mathbb{P}(X>0) \ge \frac{(\mathbb{E}X)^2}{\mathbb{E}X^2}.$$

For example, for the m balls, n bins, if $m < (1-\varepsilon)\, n\log n$, then there is an empty bin w.h.p. We have

$$\mathbb{E}X^2 = \mathbb{E}(X_1+\dots+X_n)^2 = \mathbb{E}\Big(\sum_i \underbrace{X_i^2}_{X_i} + \sum_{i\ne j} X_iX_j\Big)$$

$$= n\cdot \mathbb{E}X_1 + n(n-1)\,\mathbb{E}X_1X_2,$$

$$\mathbb{E}X_1X_2 = \mathbb{P}(X_1 = 1 = X_2) = \left(\frac{n-2}{n}\right)^m, \quad \text{so}$$

$1-\frac{1}{n} \approx e^{-\frac{1}{n}}$

$e^{-m/n} \approx e^{-(1-\varepsilon)\log n} = n^{-1+\varepsilon}$

$$\mathbb{P}(X>0) \ge \frac{n^2(1-\frac{1}{n})^{2m}}{n(1-\frac{1}{n})^m + n(n-1)(1-\frac{2}{n})^m}$$

$$\approx \frac{n^2 e^{-2m/n}}{ne^{-m/n} + n^2 e^{-2m/n}}$$

$$\approx \frac{n^2 n^{-2+2\varepsilon}}{n\cdot n^{-1+\varepsilon} + n^2 n^{-2+2\varepsilon}} = \frac{n^{2\varepsilon}}{n^\varepsilon + n^{2\varepsilon}}$$

$$= 1 - \frac{n^\varepsilon}{n^\varepsilon + n^{2\varepsilon}} = 1 - \frac{1}{1+n^\varepsilon},$$

$$\mathbb{P}(\text{there are empty bins}) \gtrsim 1 - \frac{1}{n^\varepsilon}.$$

E.g. Probabilistic method : if $X : \Omega \to \mathbb{R}$ is a r.v. s.t.

$$\mathbb{E}X > a \qquad \text{for some } a$$

then there exists $\omega \in \Omega$ s.t. $X(\omega) > a$, for otherwise

$$\forall \omega \quad X(\omega) \le a \quad \Rightarrow \quad \mathbb{E}X \le a.$$

There are m unit vectors $v_1, \dots, v_m$ in $\mathbb{R}^n$. Show that

$$\| \varepsilon_1 v_1 + \dots + \varepsilon_m v_m \| \ge \sqrt{m}$$

for some choice of signs $\varepsilon_1, \dots, \varepsilon_m \in \{-1, 1\}$.

Consider $X = \| \underbrace{\varepsilon_1 v_1 + \dots + \varepsilon_m v_m}_{\text{iid random signs}} \|^2$. We have

$$\mathbb{E}X = \mathbb{E} \langle \sum_i^m \varepsilon_i v_i, \sum_i^m \varepsilon_j v_j \rangle = \mathbb{E}\Big(\sum_i \varepsilon_i^2 \langle v_i, v_i \rangle + \sum_{i \neq j} \varepsilon_i \varepsilon_j \langle v_i, v_j \rangle \Big)$$

$$= \sum_{i=1}^m \|v_i\|^2 + \sum_{i \neq j} \langle v_i, v_j \rangle \, \mathbb{E} \varepsilon_i \varepsilon_j$$

$$= m$$

so there is a choice of $\varepsilon_1, \dots, \varepsilon_m$ s.t. $X(\varepsilon) \ge m$

(o/w $\forall \varepsilon \; X(\varepsilon) < m \Rightarrow \mathbb{E}_\varepsilon X < m$). $\square$