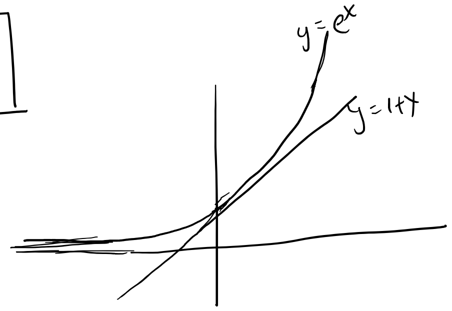
Useful Inequalities

09 February 2021 10:04

$$\boxed{1+x \leq e^{x}}$$



03/03/2021

- Cauchy-Schwarz Inequality
- Young's Inequality

Höl - der's Inequality

- Start with "online" load balancing/ Makespan

Cauchy schwarz

$$|\langle u, v\rangle| \leq \|u\|_2 \cdot \|v\|_2$$

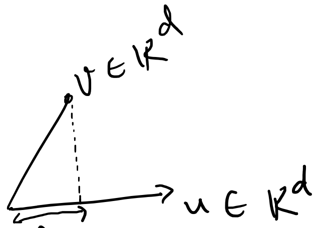
where $\|u\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_d^2}$

$$\langle u, v\rangle \equiv u \cdot v \equiv u^T v = \sum u_i v_i$$

$$\left(\sum u_i v_i\right)^2 \leq \left(\sum u_i^2\right)\left(\sum v_i^2\right)$$

Pictorial Proof

Assumed How to do Projections



$$\boxed{\frac{\langle u, v\rangle}{\|u\|} \leq \|v\|}$$

Young's Inequality :-

if p and $q \geq 0$ are st

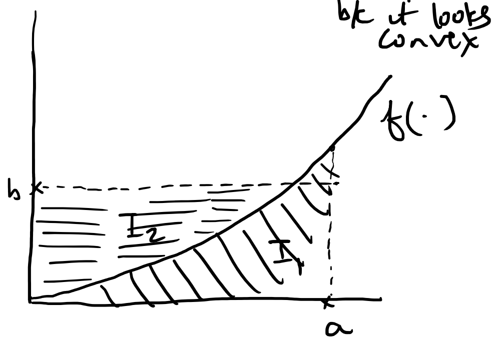
$$\frac{1}{p} + \frac{1}{q} = 1 \quad \text{then}$$

for all $a \geq 0$, $b \geq 0$,

$$\boxed{ab \leq \frac{a^p}{p} + \frac{b^q}{q}}$$

Proof

Let $f(x) = x^{p-1}$



In geometric view,

$$ab = \text{Area of Rectangle}.$$

$$\leq I_1 + I_2.$$

Apply to $f(x) = x^{p-1}$

$$I_1 = \left.\frac{x^p}{p}\right]_0^a = \frac{a^p}{p}.$$

I_2 = Integral of the inverse function

$$y = x^{p-1}$$

$$y^{\frac{1}{p-1}} = x$$

$$f^{-1}(y) = y^{\frac{1}{p-1}}$$

$$I_2 = \left. \frac{y^{\frac{p}{p-1}}}{\frac{p}{p-1}} \right|_0^b$$

$$= \frac{b^{\frac{p}{p-1}}}{\frac{p}{p-1}} = \frac{b^q}{q}.$$

Recall

$\frac{1}{p} + \frac{1}{q} = 1 \Rightarrow q$ is exactly $\frac{p}{p-1}$

Hence we get $\boxed{ab \le \frac{a^p}{p} + \frac{b^q}{q}}$

For $p = q = 2$, this is famous

$2ab \le a^2 + b^2$ inequality

Hölder's Inequality

Can think of it as a generalization of Cauchy-Schwarz for l_p norm where $p \neq 2$.

Any vector $v \in \mathbb{R}^d$ has many notions of how "long" it is.

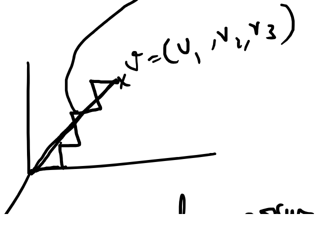
$\downarrow$

Each is called a norm.

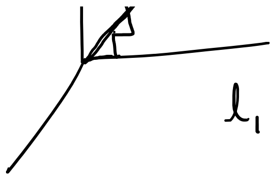
$v = (v_1, v_2, \ldots, v_d)$

l_2 norm $\quad \|v\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_d^2}$

"distance as the bird would fly"



l_1 norm (Manhattan Norm)



l_1 norm (Manhattan Norm)

$= |v_1| + |v_2| + \dots + |v_d|$

l_p Norm $= \left(|x_1|^p + |x_2|^p + \dots + |x_d|^p \right)^{1/p}$

Any function

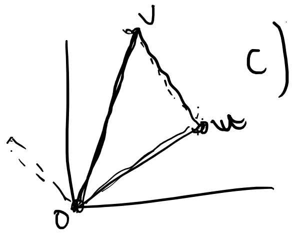
$f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ is a NORM iff

a) $f(\alpha \cdot v) = |\alpha| f(v) \quad \forall \alpha \geq 0, \ \forall v \in \mathbb{R}^d$

b) Δ^{le} inequality is true.

(ie) $\forall u \in \mathbb{R}^d, v \in \mathbb{R}^d$

$f(v) \leq \underline{f(u)} + f(v-u)$



c). $f(v) = 0$ iff $v = 0$

EXERCISE

Verify that l_1, l_2 & l_p are indeed NORMS.

Hölders Inequality

Extension of Cauchy Schwarz for l_p NORMS.

In general, if

$v = (v_1, v_2, \dots v_d)$

$u = (u_1, u_2, \ldots u_d)$

Hölder

$$\boxed{\sum |u_i v_i| \leq \|u\|_p \cdot \|v\|_q}$$

if $\frac{1}{p} + \frac{1}{q} = 1$.

[One of the motivations for calling $\|\cdot\|_q$ as the "dual norm" of $\|\cdot\|_p$

whenever $q = \frac{p}{p-1}$.

Idea: let's try to simplify the problem without loss of generality

Can we assume that $\|u\|_p = 1$ & $\|v\|_q = 1$?

Yes, because

let $\hat{u}_i = \frac{u_i}{\|u\|_p}$, $\hat{v}_i = \frac{v_i}{\|v\|_q}$

Now, $\|\hat{u}\|_p = \|\hat{v}\|_q = 1$ (by scaling property)

$\Rightarrow$ $\boxed{\text{If we show } \sum_i |\hat{u}_i \hat{v}_i| \leq 1}$ (*)

$$\Rightarrow \sum \left| \frac{u_i}{\|u\|_p} \cdot \frac{v_i}{\|v\|_q} \right| \leq 1$$

$$\Rightarrow \sum |u_i v_i| \leq \|u\|_p \|v\|_q. \quad \blacksquare$$

Remains to show ($*$)

Let's apply Young's Inequality inside each term

$$|\hat{u}_i \hat{v}_i| \leq \frac{|\hat{u}_i|^p}{p} + \frac{|\hat{v}_i|^q}{q}$$

Sum over i

$$\sum |\hat{u}_i \hat{v}_i| \leq \frac{\sum_i |\hat{u}_i|^p}{p} + \frac{\sum_i |\hat{v}_i|^q}{q}$$

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

$$= 1 .$$ (dual norm)

Stirling's Approximations

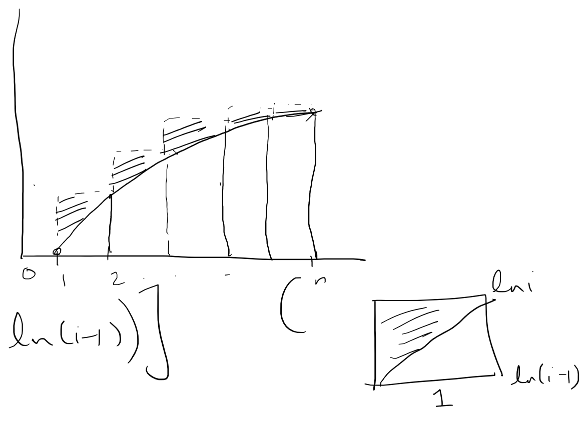
01 February 2021 09:38

I $\boxed{n! \approx \sqrt{n} \cdot \left(\frac{n}{e}\right)^n}$

How ?

$$Q = \ln(n!) = \sum_{i=1}^{n} \ln n$$

$$I = \int_1^n \ln x \, dx$$



$$I - Q = \text{shaded region} \approx \frac{1}{2}\left[\sum \left(\ln i - \ln(i-1)\right)\right]$$

$$\approx \frac{\ln n}{2}$$

$$\therefore \quad Q = I - \ln\sqrt{n}$$

$$= \left[x\ln x - x\right]_1^n - \ln\sqrt{n}$$

$$= n\ln n - n - \ln\sqrt{n}$$

$$= \ln\left(\sqrt{n} \cdot \left(\frac{n}{e}\right)^n\right)$$

$$\Rightarrow n! \approx \sqrt{n} \cdot \left(\frac{n}{e}\right)^n$$

II $\boxed{\binom{n}{k} \approx \left(\frac{ne}{k}\right)^k}$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{\sqrt{n} \; n^n \; \cancel{e^k} \; \cancel{e^{n-k}}}{\sqrt{k} \; \cancel{e^n} \; k^k \sqrt{n-k} \; (n-k)^{n-k}}$$

$$\approx \left(\frac{n}{k}\right)^k \left[1 + \frac{k}{n-k}\right]^{n-k}$$

$$\approx \left(\frac{ne}{k}\right)^k \quad \boxtimes$$

Gaussian Tails

02 February 2021 14:37

$$Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2}\,dx$$

$$Q(t) \le \frac{1}{\sqrt{2\pi}} \int_t^\infty \left(\frac{x}{t}\right) e^{-x^2/2}\,dx$$

$$= \frac{1}{t\cdot\sqrt{2\pi}} \int_t^\infty x e^{-x^2/2}\,dx = \frac{1}{t\cdot\sqrt{2\pi}}\, e^{-t^2/2}$$

$$= \frac{f(t)}{t}.$$

On the other hand

$$\left(1+\frac{1}{t^2}\right) Q(t) \ge \frac{1}{\sqrt{2\pi}} \int_t^\infty \left(1+\frac{1}{x^2}\right) e^{-\frac{x^2}{2}}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \left[-\frac{e^{-x^2/2}}{x} \right]_t^\infty = \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}$$

$$\Rightarrow Q(t) \ge \frac{t}{1+t^2}\cdot f(t)$$

$$= \frac{1}{t}\left[1 - \frac{1}{1+t^2}\right] f(t)$$

$$\ge \left(\frac{1}{t} - \frac{1}{t^3}\right) f(t)$$

Hence

$$\boxed{\left(\frac{1}{t} - \frac{1}{t^3}\right) f(t) \le Q(t) \le \frac{1}{t}\cdot f(t)}$$

- Broad class of Problems can be modeled with linear programming

- n real valued variables $x_1, x_2, \dots, x_n$
  - m linear constraints over those variables

one linear objective function over them.

either maximize (or) minimize Obj. fn. subject to all constraints

Useful in theory & practice

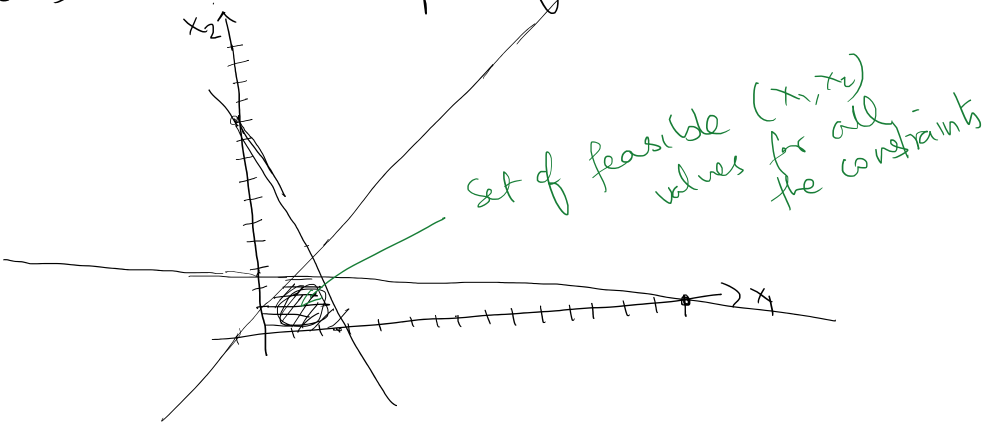
[Matousek
Understanding and Using Linear Programming]

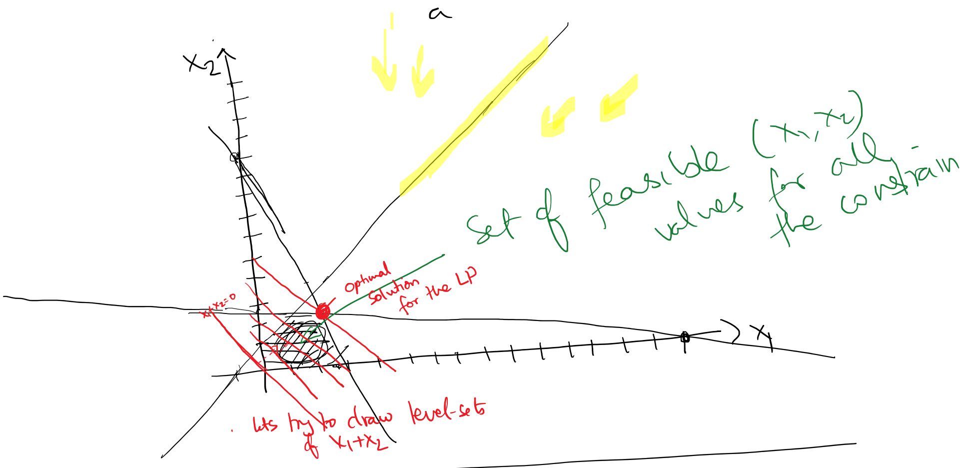
Max $x_1 + x_2$

$$x_2 - x_1 \leq 1$$
$$x_1 + 6x_2 \leq 15$$
$$4x_1 - x_2 \leq 10$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$

{LPs don't allow strict Inequalities}

How does the solution space of this look like?





General Form

$$\boxed{\begin{aligned} \min\ & c^T x \\ a_1^T x &\geq b_1 \\ a_2^T x &\geq b_2 \\ &\vdots \\ a_m^T x &\geq b_m \end{aligned}}$$

$c^T x = \sum c_i x_i$
$= \langle c, x \rangle$
$= c \cdot x$

$c^T x = (c_1 c_2 \cdots c_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

$\sum c_i x_i$

$\Downarrow\Uparrow$

$$\boxed{\begin{aligned} \min\ & c^T x \\ Ax &\geq b \end{aligned}}$$

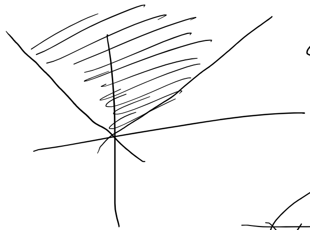
typically $m \geq n$

$A = m \times n$ matrix
rows of A correspond to the constraints

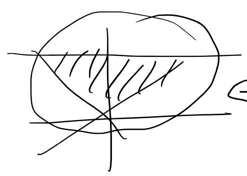
What does the soln space look like?
Intersection of 'm' halfspaces

Polyhedron (Intersection of finitely many half spaces)

Polytope
Polyhedron which is also bounded



Polyhedron but not a polytope



polytope

Feasible set for any LP is

Feasible set for any LP is
a polyhedron
How do we characterize Optimal Solutions of
LP ?

5th Feb 2021

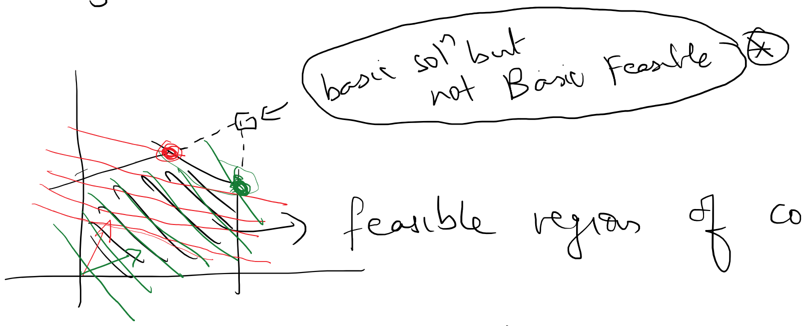
$x_1, x_2, \dots x_n \in \mathbb{R}$ are variables

$$\left\{ \begin{array}{ll} \max & c^T x \\ & Ax \le b \end{array} \right\}$$

Optionally can separate out
$x_i \geq 0$ type
"non-negativity"
constraints
If they are present

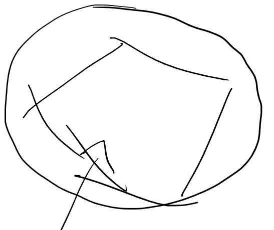
Why Study LPs.
① Very general, can capture variety of problem
② You can solve them efficiently
In theory & in practice

Why are we able to solve them so effectively ?



feasible region of constraints,

For this lecture (and most of the course, & most applications)
Our feasible sets will be bounded



Can be encapsulated inside some n-dimensional ball of finite radius

POLYTOPES
(bounded intersection of finitely many half-spaces)

LPs are nice to solve over polytopes because optimal solutions always occur at "corner points"

Q: How do we characterize a corner point?

Choose n out the m constraints and solve them @ equality

$$\max c^T x$$
$$\underset{m}{\updownarrow} \; Ax \le b$$

$A \in \mathbb{R}^{m \times n}$
$b \in \mathbb{R}^{m}$
$c \in \mathbb{R}^{n}$

lets assume $m \geq n$

$$A_S \cdot x = b_S$$

notice the equality

$S \subseteq [m]$ of size n

Sps $\det(A_S) \neq 0$

$$\boxed{\text{then } \mathrm{sol}^n \text{ is } x^{(S)} = A_S^{-1} b_S}$$

Such solution are called BASIC SOLUTIONS

Now, it might not satisfy the other constraints in $[m] \setminus S$

If $x^{(S)}$ additionally satisfies

all other constraints $\longrightarrow$ $A_{[m]\setminus S} \, x^{(S)} \le b_{[m]\setminus S}$

$x^{(S)}$ is called a basic Feasible Solution

See ⊛ above for example of

See ⊛ above for example of basic sol^n which is not BFS.

{At most $\binom{m}{n}$ many BFS, can check in finite time & optimize}

There are algos which find the Optimal BFS in poly time !!

<u>Geometric Views of "Corner Points"</u>

① VERTEX OF POLYTOPE P

$x \in P$ is a "vertex" if

$\exists$ objective function $c \in \mathbb{R}^n$

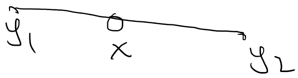
st $c^T x > c^T y \quad \forall y \in P, y \neq P$

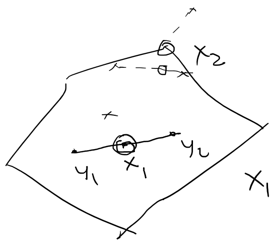
② <u>Extreme Points</u> of Polytope P

$x \in P$ is an <u>extreme point</u>

iff $\nexists$ $y_1, y_2 \in P$, $y_1 \neq y_2$ st

$x = \alpha y_1 + (1-\alpha) y_2$ for some $\alpha \in [0,1]$





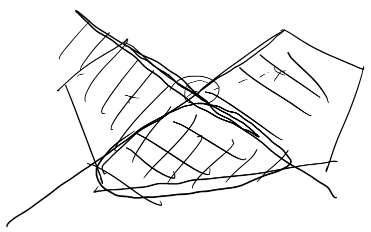
x_1 is not an extreme point

x_2 is an extreme point

$$\boxed{\begin{array}{l}\forall \text{ polytope} \\ \text{Extreme Points} = \text{Vertices} = \text{BFS}^s\end{array}}$$

Very useful THEOREM

Very useful THEOREM



← Such polygons will not arise as the feasible region of any LP

THM

Intersection of finitely many halfspaces is convex

$\Rightarrow$ polytope P is convex

DUALITY

Useful way to understand optimal solns of LPS without "optimizing the LP".

$$\text{Max} \quad 2x_1 + 3x_2$$

$$4x_1 + 8x_2 \le 12 \quad (1)$$

$$3x_1 + 2x_2 \le 4 \quad (2)$$

$$2x_1 + x_2 \le 3 \quad (3)$$

$$x_1 \geq 0$$

$$x_2 \geq 0$$

Can we get good "upper bounds" on the optimal value of the LP without solving it?

Due to $x_1 \geq 0$ & $x_2 \geq 0$

$$2x_1 + 3x_2 \le 2x_1 + 4x_2 = \frac{1}{2}(4x_1 + 8x_2) \le 6$$

From constr (1)

$\Rightarrow$ Optimal Soln ≤ 6

Can we do better?

(1) + (3) gives

$$6x_1 + 9x_2 \le 15$$

$$\Rightarrow \boxed{2x_1 + 3x_2 \le 5}$$

$\Rightarrow \boxed{2x_1 + 3x_2 \leq 5}$

equivalently,

$\frac{1}{3}(1) + \frac{1}{3}(3)$

Since we're looking for upper bounds, we can try "dominating" the objective fn c by non-negative linear combination of the constraints, to get a best upper bound.

$$
\begin{array}{lll}
\max & c^T x & \\
y_1 & a_1^T x & \leq b_1 \\
y_2 & a_2^T x & \leq b_2 \\
\vdots & \vdots & \\
y_m & a_m^T x & \leq b_m \\
 & x_1 \geq 0 & \\
 & x_2 \geq 0 & \\
 & \vdots & \\
 & x_n \geq 0 &
\end{array}
$$

best such "upper bound" can be found by a linear program in itself.

Seeking multiplying factors $y_1 \cdots y_m \geq 0$

Coefficient of x_j in this combination $= \sum_{i=1}^{m} y_i \, a_{ij}$

Want $\underbrace{c_j \leq \sum_{i=1}^{m} y_i \, a_{ij}}$ $\forall 1 \leq j \leq n$

$\Rightarrow \sum c_j x_j$ (for any feasible sol^n x)

$\leq \sum b_i y_i$

---

gives us DUAL program

$\min \sum b_i y_i = b^T y$

$$
\boxed{\begin{array}{l} \max\ c^T x \\ Ax \leq b \\ x \geq 0 \end{array}}
\ \underset{\text{dualize}}{\overset{\text{dualize}}{\rightleftarrows}}\
\boxed{\begin{array}{l} \min\ b^T y \\ A^T y \geq c \\ y \geq 0 \end{array}}
$$

$Ax \le b$
$x \ge 0$

dualize

$Ay \ge c$
$y \ge 0$

Dual (Dual) = Primal

WEAK DUALITY

THM: If x^* is optimal sol^n for primal & y is <u>any</u> <u>feasible</u> sol^n for Dual

$$\boxed{c^T x^* \le b^T y}$$

{In form discussed $x \ge 0$ & we had $\le$ constraints}

Mechanical dual generator in most general form.

$$P = \max c^T x$$

$$a_i^T x \le b_i \quad \forall i \in I_1$$

$$a_i^T x = b_i \quad \forall i \in I_2$$

$$x_j \ge 0 \quad \forall j \in J_1$$

$$x_j \in \mathbb{R} \quad \forall j \in J_2$$

i^{th} row of A

$$D = \min b^T y$$

$$y_i \ge 0 \quad \forall i \in I_1$$

$$y_i \in \mathbb{R} \quad \forall i \in I_2$$

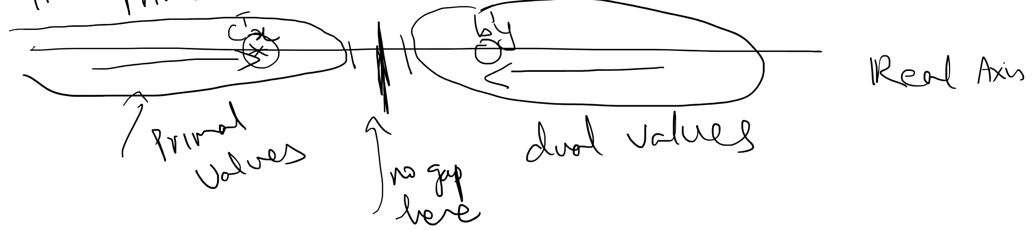
$$A_j^T y \le c_j \quad \forall j \in J_1$$

$$A_j^T y = c_j \quad \forall j \in J_2$$

j^{th} column of A

<u>Todo</u>: Practice duals for many LPs.

IF Primal is maximization & Dual is minimization



STRONG DUALITY THEOREM

Given P & D, One of 4 cases can occur :

1. Either P & D are Infeasible
2. Either P is UNBOUNDED $\Rightarrow$ D is Infeasible
3. Either D is UNBOUNDED $\Rightarrow$ P is Infeasible
4. Both feasible & optimal values equal (ie) $\boxed{c^T x^* = b^T y^*}$

Proof of Strong Duality

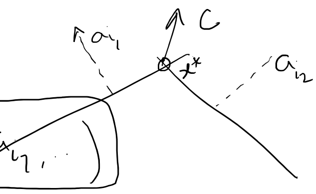
let x^* be primal optimal for

$$\max \ c^T x$$
$$Ax \le b$$

and let I be indices which satisfy tight inequality

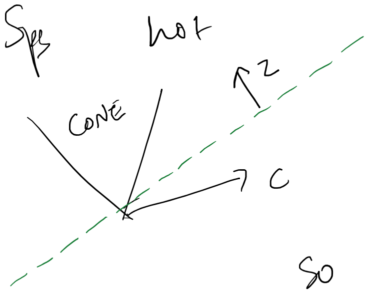
(ie) $a_i^T x^* = b_i \ \forall i \in I$

CLAIM :-



& $\boxed{c \in \text{Cone}(a_1, a_2, \dots)}$

Sps not



$\exists$ hyperplane st

$z^T \cdot \text{CONE} \le 0$ &

$z^T c > 0$

So we can perturb

$x^* = x^* + \epsilon \cdot z$

and violate optimality of x^*

Being in cone $\Rightarrow$

$c = \sum y_i^* a_i$

set $y_i^* = 0 \ \forall i \notin I$ and so

$$A^T y^* = c$$

$$b^T y^* = \sum b_i y_i^* = \sum_{i \in I} y_i^* a_i^T x^*$$

$$= \langle x^*, \sum y_i^* a_i \rangle$$

$$= c^T x^* \quad \boxtimes$$

Some Useful Probabilistic Inequalities
The study of random variables' behaviour

Markov's Inequality

If X is a non-negative random variable,
the $Pr[X \geq t\,E[X]] \leq \frac{1}{t}$

equivalently
$$Pr[X \geq v] \leq \frac{E[X]}{v}$$

Recall
defn of $E[X] = \sum_{v} v\, Pr(X = v)$ or
$= \sum_{\substack{\text{all possible} \\ \text{outcomes } \omega}} val(X_\omega) \cdot Pr(\omega)$

also recall linearity of Expectation
$$E[X_1 + X_2] = E[X_1] + E[X_2]$$
regardless of (in)dependence of X_1 & X_2

How useful is Markov's Inequality

Sps we toss 1000 coins independently
each H wp 1/2 &
T wp 1/2

$X_i = 0$ if Tail
$\quad\;\; 1$ if Head for i^{th} coin

$$E[X_i] = \frac{1}{2}$$
$$E[X] = E[\Sigma X_i] = 500$$

$\boxed{\text{How likely / unlikely to get in excess of 750 }\cancel{\text{heads}}}$

Markov's Inequality gives

$$\Pr(X \geq 750) \leq \frac{500}{750} = \frac{2}{3}.$$

Loose estimate
(does not make use of fact that all X_i's are independent at all)

<u>Proof</u> of Markov's Inequality

$$\Pr[X \geq V] \leq \frac{E[X]}{V}$$

Q. Can we make use of the independence of X_i's to get better estimates.

<u>DETOUR</u>

$E[X] = V$

$E[cX]$ for c constant $= cV$

$E[C^2]$

but $E[X^2] \neq V^2$

Eg: $X = \begin{cases} -1 & \text{w/p } 1/2 \\ 1 & \text{wp } 1/2 \end{cases}$ $E[X] = 0$ but $E[X^2] = 1$

DEFINE

$$Var(X) = E\left[(X - E[X])^2\right]$$

$X = \begin{cases} 0 & \text{wp } 1/2 \\ 1 & \text{wp } 1/2 \end{cases}$ $E[X] = \frac{1}{2}$

$Var(X) = \frac{1}{4}$

$X - E[X] = \begin{cases} -1/2 & \text{wp } 1/2 \\ +1/2 & \text{wp } 1/2 \end{cases}$ $E[(X - E[X])^2] = \frac{1}{4}$

$$Var(X) = E\left[X^2 - 2\underbrace{X E[X]} + E[X]^2\right]$$

linearity of expectation $= E[X^2] - 2E[X] \cdot E[X] + E[X]^2$

$$= E[X^2] - E[X]^2$$

In 0/1 example $E[X^2] = 1/2$
$E[X]^2 = 1/4$

Let's go back to the coins example
(recall 1000 coins, trying to understand $Pr(\geq 750 \text{ heads})$)

$$X = \sum_{i=1}^{1000} X_i$$

let's make life easy a bit and consider slightly changed random variables

$Y_i = -1$ if TAIL for ith coin

$= 1$ if HEAD

$$E[Y_i] = 0$$

$$Y = \sum Y_i$$

Q: How are X & Y related?

A: $Y_i = 2(X_i - \frac{1}{2}) = 2X_i - 1$

$$\boxed{Y = 2X - n}$$

Trying to study deviations for $X \rightleftharpoons$ deviations for Y

e.g. $E[Y] = 0$

$$\Pr(X \geq 750) = \Pr(Y \geq 500)$$

$$\mathrm{Var}(Y) = E[(Y - E[Y])^2]$$

$$= E[Y^2]$$

$$= E[(\sum Y_i)^2]$$

$$= E[\sum Y_i^2 + \sum_{i \neq j} Y_i Y_j]$$

$$= \sum E[Y_i^2] + \sum_{i \neq j} \underbrace{E[Y_i Y_j]}$$

$$\left\{ \begin{array}{l} \Pr(Y = y \mid X = x) \\ = \Pr(Y = y) \\ \forall (x, y) \end{array} \right.$$

If X & Y are "independent" random variables, $E[XY] = E[X] \cdot E[Y]$.

$$\sum^{n} E[Y_i^2] = n.$$

$$\boxed{Var(Y) = \sum_{i=1}^{n} E[Y_i^2] = n}$$

$$Pr(Y \geq 500) \leq Pr(|Y| \geq 500) = Pr(Y^2 \geq 500^2)$$

$$\leq \frac{E[Y^2]}{500^2} \quad \text{MARKOV}$$

equal to Var(Y)

$$= \frac{1000}{500 \cdot 500}$$

$$= \frac{2}{500}$$

$$= \frac{1}{250}$$

by using "PAIRWISE independence" of X_i & X_j (Y_i & Y_j) we get much better bounds.

Chebyshev's Inequality

If X is any random variable,

$$E[X] = \mu$$

$$Var[X] = E[X^2] - E[X]^2 = \sigma^2$$

Then

$$Pr[\,|X - \mu| \geq t\sigma\,] \leq \frac{1}{t^2}$$

Proof: use Markov on R.V $Y = (X-\mu)^2$

Why stop here?

These random variables (for the coins)
are not just pairwise independent
Any 3 of them are independent
In fact, Any subset of them are "

Can we use this fact to
get better estimates?

Try to use
Markov's on
Y^{2k} for some good choice of k

$Y = \sum Y_i$ $\quad E[Y_i] = 0 \quad \forall i$

Y_i's are independent

Each Y_i : -1 wp $1/2$, 1 wp $1/2$ $\quad E[Y] = 0$

$$\Pr(Y \geq 500) \leq \Pr(Y^{2k} \geq 500^{2k})$$
$$\leq \frac{E[Y^{2k}]}{500^{2k}}$$

Try to understand $E[Y^{2k}]$

$$= E\left[(\sum Y_i)^{2k}\right]$$

" be) dominated by

will be dominated by

$$\sum \left(Y_1^2 Y_2^2 \cdots Y_k^2\right)$$

$$\binom{n}{k} \cdot \frac{(2k)!}{\underbrace{2 \cdot 2 \cdots 2}_{2^k}}$$

$E[Y^{2k}]$ will be approximately $\binom{n}{k} \frac{(2k)!}{2^k}$ $\ast$

Stirling's Approximation For Factorials

$$\binom{n}{k} \approx \left(\frac{ne}{k}\right)^k$$

$$n! \approx \sqrt{n} \cdot \left(\frac{n}{e}\right)^n$$

$\ast$ gives $\frac{n^k \cdot e^k}{k^k} \cdot \frac{(2k)^{2k}}{e^{2k} \cdot 2^k}$

$$\frac{n^k \cdot \cancel{e^k}}{\cancel{k^k}} \cdot \frac{2^{2k} \cdot k^{2k}}{e^{2k} \cdot \cancel{2^k}}$$

$$= \left(\frac{2nk}{e}\right)^k$$

$$\Pr(Y \geq t \cdot \sqrt{n}) \leq \frac{E[Y^{2k}]}{(t\sqrt{n})^{2k}}$$

$$(t\sqrt{n})^{2k}$$

$$= \frac{\left(\frac{2nk}{e}\right)^k}{t^{2k} \cdot n^k}$$

$$= \left(\frac{2nk}{e n t^2}\right)^k$$

$$= \left(\frac{2k}{e t^2}\right)^k$$

I am free to choose k, we'll optimize to minimize RHS.

We'll just set $k = t^2$

$$\Pr(Y \geq t\sqrt{n}) \leq \left(\frac{2}{e}\right)^{t^2}$$

SUMMARY

Markov

$$\Pr(X \geq t\, E[X]) \leq \frac{1}{t}$$

Chebyshev

$$\Pr(|X - E(X)| \geq t\sigma) \leq \frac{1}{t^2}$$

Chernoff $\Pr(|X - E(X)| \geq t\sigma) \leq \exp(-t^2)$

X is decomposable as sum of indep RVs

$$\Pr(750 \text{ heads}) \leq .0000$$

$Pr(\geq 750 \text{ heads}) \leq \cdot 0000$

SUMMARY

If X is non-negative R.V

Markov's Inequality

$$Pr[X \geq \lambda E[X]] \leq \frac{1}{\lambda}$$ } Multiplicative deviation from than mean $E[X]$

Chebyshev's Inequality

X is any R.V, with Expectation $E[X]$ &
Variance $E[X^2] - E[X]^2 = \sigma^2$

$$Pr[|X - E[X]| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}$$

Chernoff Bounds

if X is the SUM of independent & bounded random variables,

the X concentrates sharply around its mean

Popular Forms

1) if $X = \sum X_i$ each $X_i \in [0,1]$

a) $$Pr[X \leq E[X](1-\delta)] \leq e^{-\frac{E[X]\delta^2}{2}} \quad \forall\, 0 \leq \delta \leq 1$$

b) $$Pr[X \geq E[X](1+\delta)] \leq e^{-\frac{E[X]\delta^2}{2+\delta}} \quad \forall\, \delta \geq 0$$

(2) if $X = \sum X_i$, where X_i's are independent random variables with $|X_i| \le M$

and $Var(X) = E[X^2] - E[X]^2 \le \sigma^2$

$$\Pr\left[|X - E[X]| \geq T\right] \leq \exp\left(\frac{-T^2}{\sigma^2 + \frac{MT}{3}}\right)$$

BERNSTEIN'S INEQUALITY

for some intuition,
sps we plug in $T = \lambda\sigma$
& we ignore $\frac{MT}{3}$ (often this will be small)

$$\Pr\left(|X - E[X]| \geq \lambda\sigma\right) \overset{\approx}{\leq} \exp\left(-\lambda^2\right)$$

TODO

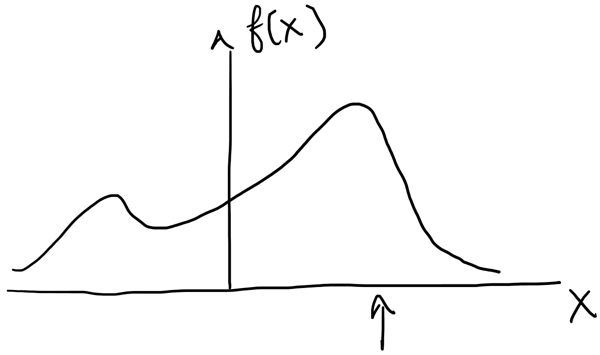
Try these inequalities to see what they yield for the coins problem

(1000 coins, independent, what is probability of ≥ 750 heads)

More on probability will be later — Gaussian random variables.

Question

X is a random variable X over real values and has some distribution

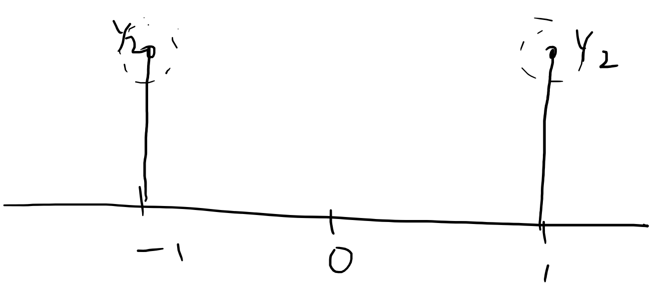


Suppose we consider $Y = X_1 + X_2$ where X_1 and X_2 are 2 independent copies of X.

What does the distribution of Y look like?

EXAMPLE

$$X = \begin{cases} +1 & \text{w.p } 1/2 \\ -1 & \text{w.p } 1/2 \end{cases}$$



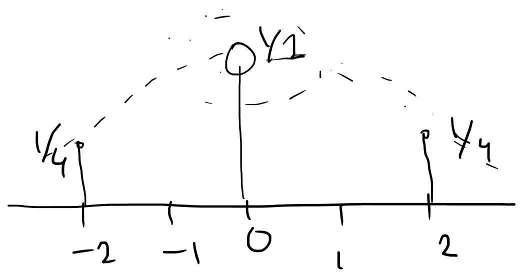
← bi-modal distribution (2 peaks).

$Y = X_1 + X_2$

$Pr(Y=0) = \frac{1}{2}$

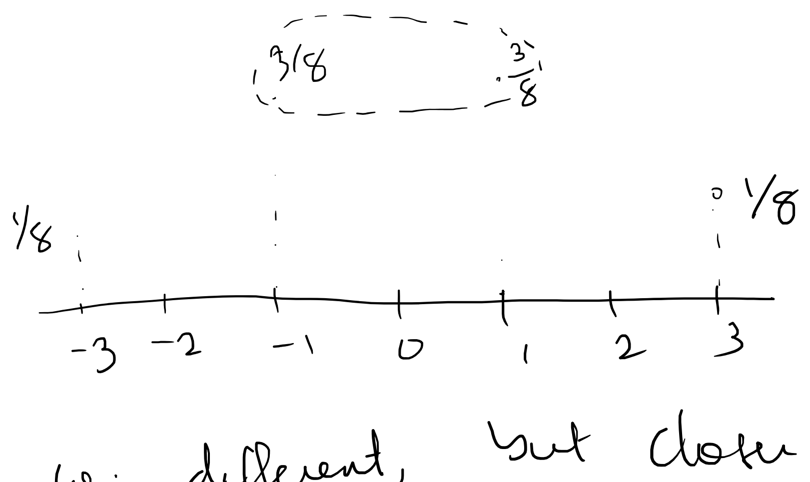
$Pr(Y=1) = 0$

$Pr(Y=2) = \frac{1}{4}$



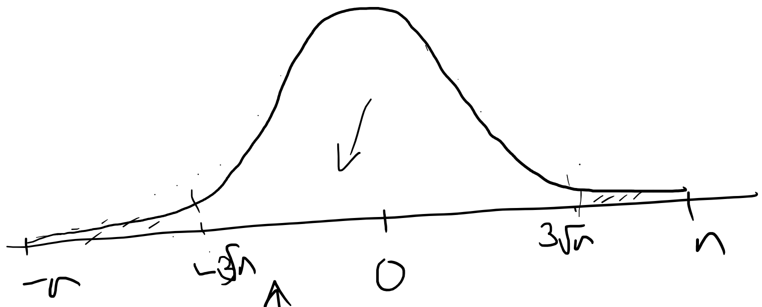
[distribution of Y is different from that of X.]

$Y = X_1 + X_2 + X_3$ (similarly)



looks different, but closer to having one peak

Eventually, you add enough times, you'll end up with the following type of distribution $Y = X_1 + X_2 + \cdots + X_n$



Normal distribution

Moreover

This behaviour is not just for ± 1 R.V's.

Central Limit Theorem

Let X be any random variable.

Let $Y = X_1 + X_2 + \cdots + X_n$ be n copies of X (independent).

Then Y is almost distributed like a "Gaussian Distribution".

Y will have mean $E[Y] = n \cdot E[X]$

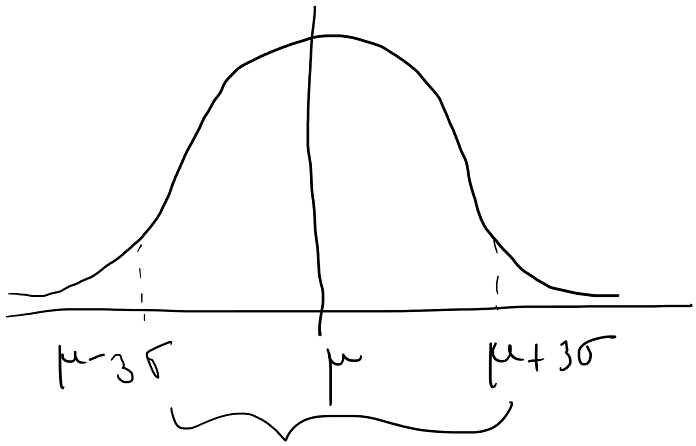
and variance $Var[Y] = \sum Var[X_i] = n \cdot Var[X]$.

DEFN:

$X \sim N(\mu, \sigma^2)$ is a gaussian

$X \sim N(\mu, \sigma^2)$ is a gaussian random variable with mean μ and variance σ^2 iff it has the following probability density function :-

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(t-\mu)^2}{2\sigma^2} \right)$$



Most of the probability ($> 95\%$) is in the interval

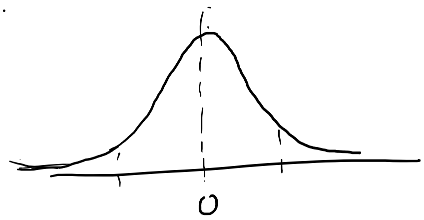
$$[\mu \pm 3\sigma].$$

Standard Normal Distribution

$X \sim N(0,1)$ Mean 0
Variance 1

$-t^2/2$

$$f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$



These $\exp\left(-\frac{t^2}{\sigma^2}\right)$ type bound occurred in Chernoff bounds also

[Gaussians are deeply interconnected with these inequalities].

For any gaussian random variable

$$X = N(0, 1).$$

$$E[X] = \int_{-\infty}^{\infty} t \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \left[e^{-t^2/2} \right]_{-\infty}^{\infty}$$

$$= 0$$

$$\mathrm{Var}[X] = E[(X-\mu)^2] = E[X^2]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2/2}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty} t^2 e^{-t^2/2} \, dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underset{u}{t} \, \Big(\underbrace{t e^{-t^2/2} \, dt}_{dv} \Big)$$

$$= \frac{1}{\sqrt{2\pi}} \left[uv \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v \, du \right]$$

$$= \frac{1}{\sqrt{2\pi}} \left[t \cdot e^{-t^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-t^2/2} \, dt \right]$$

$$= \frac{1}{\sqrt{2\pi}} \left[0 + \int_{-\infty}^{\infty} e^{-t^2/2} \, dt \right]$$

$$= \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt} = \underline{1}$$

because - f is a pdf.

↓

also (More importantly) because

$$\boxed{\int_{-\infty}^{\infty} e^{-t^2/2} \, dt = \sqrt{2\pi}}$$

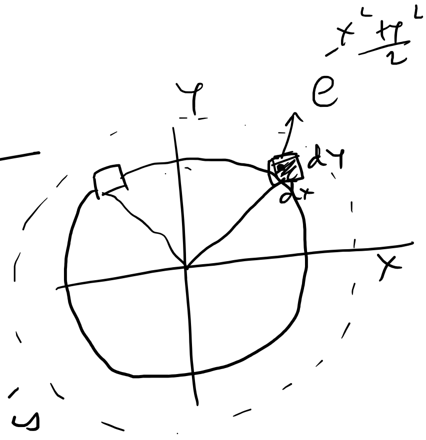
$$\boxed{\int_{-\infty}^{\infty} e^{-t^2/2}\, dt = \sqrt{2\pi}}$$

Proof

let $I = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$

$$I = \int_{-\infty}^{\infty} e^{-y^2/2}\, dy$$

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\, dx\, dy.$$



because it is spherically symmetric, (ie) the value of the integrand is invariant under rotation due to its form of $\exp\left(\frac{-(x^2+y^2)}{2}\right)$

We can use polar coordinates ☺

$$I^2 = \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} e^{-\frac{r^2}{2}}\, r\, dr\, d\theta$$

$$I = \int_{r=0} \int_{\theta=0} e$$

$$= 2\pi \int_{r=0}^{\infty} r e^{-r^2/2} \, dr$$

$$= 2\pi \left[-e^{-r^2/2} \right]_0^{\infty}$$

$$= 2\pi .$$

$$\Rightarrow \boxed{I = \sqrt{2\pi}}$$

- We have understood one gaussian well.
- what about sums of independent gaussians ?

$X_1 = N(0,1)$ and independent
$X_2 = N(0,1)$
what will the distribution of $X_1 + X_2$ look like ?

let $Y = X_1 + X_2$

$$E[Y] = E[X_1] + E(X_2] = 0.$$

$\mathrm{Var}[Y] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2] = 2.$

Distribution of Y ?

↓

Y will also be gaussian !

— $N(0, 2)$

INTUITION

① (CLT intuition

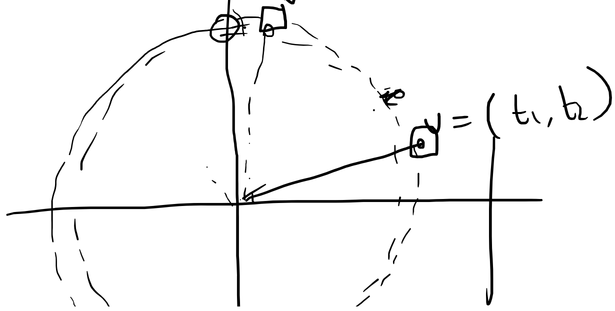
X_1 can be thought of as sum of many small RVs)

X_2 similarly

$\Rightarrow$ Y has gaussian form, with mean 0 and Var $= 2$.

INTUITION ②

lets look at joint distribution of (X_1, X_2)



$$f_{X_1,X_2}(t_1,t_2) = \frac{1}{2\pi}\exp\left(-\frac{t_1^2+t_2^2}{2}\right)$$

Spherical symmetry.

Joint distribution of (X_1, X_2) is the same as joint distribution of

$$\left(\frac{X_1+X_2}{\sqrt{2}}, \frac{X_1-X_2}{\sqrt{2}}\right)$$

rotation of (X_1, X_2) by 45°

$\Rightarrow$ distribution of $\left(\frac{X_1+X_2}{\sqrt{2}}\right)$ is identical to that of X_1

$= N(0,1)$

$\Rightarrow$ $\mathrm{dist}(X_1+X_2) = \sqrt{2}\,N(0,1)$

$= N(0,2)$.

More generally

More generally

$X_1, X_2, X_3, \ldots X_d$ are all $N(0,1)$

then

$Y = \sum a_i X_i$ is distribution as $\boxed{N(0, \sum a_i^2)}$

for scalars $a_1, a_2, \ldots, a_d$.

Given a gaussian $X = N(0,1)$

for $v \geq 0$ $\Pr(X \geq v)$ is "tail estimate"

$$\Pr(|X| \geq v) = 2 \cdot \underbrace{\int_v^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt}$$

$$2\left(\frac{1}{v} - \frac{1}{v^3}\right) f(v) \leq \Pr(|X| \geq v) \leq \frac{2}{v} f(v)$$

pdf @ v.
