KEIF CH 1 (also, cf CH 2 easy Reif)

COMPUTING & USING PROBABILITIES (fast)

SPS M THINGS CAN HAPPEN

i LABELS WHICH

ex 3 SIDED DIE (I know most have 6, but the last 3 don't teach anything)

$i = 1, 2, 3 \qquad M = 3 \qquad u_i = \#$ ON DIE :

i	u_i
1	5
2	7
3	9

MEASURE PROBS: (don't know in advance)

THROW $N = 100$ TIMES: {APPROX TO $N = \infty$}

{in practice, enough times so that all things can happen a lot}

u	# u's
5	20
7	50
9	30

(u ≡ "RANDOM VARIABLE")

Idea is probability defn: ≡ fraction of times occurs after many trials

$$P(5) = \frac{20}{100} = 0.2 \quad \text{(ie 20\%)}$$

basic idea; can use to derive some simple rules for manip. P's

$$P(7) = \frac{50}{100} = 0.5$$

$$P(9) = \frac{30}{100} = 0.3$$

("PROB. DISTRIBUTION")

⇒ EVERYTHING CAN SAY ABOUT u

⇒ CAN ANSWER ANY QUESTION w/ $P(i)$

---

(1) NORMALIZATION:

$$\sum_{i=1}^{M} P(u_i) = 1$$

≡ PROB. ~~SOMETHING~~ HAPPENED (ie $\frac{20 + 50 + 30}{100}$)

⇒ ONLY NEED REL. PROBS

ex $P(A) = 2\,P(B) = 4\,P(C) \Rightarrow \cancel{P(A) = \frac{4}{7}} \quad P(B) = \frac{2}{7} \quad P(C) = \frac{1}{7}$

(1) ADDING PROBS:

- ONE MEAS.
- EITHER / OR

ex PROB 5 OR 7 (ie NOT 9)

$N = 100$

OCCURRENCES: $20 + 50$

$$P(5 \text{ or } 7) = \frac{20+50}{100} = \frac{20}{100} + \frac{50}{100} = P(5) + P(7) = 0.7$$

$\Rightarrow$ LESS RESTRICTIVE, P. INCR.

(2) NORMALIZATION:

PROB ANYTHING HAPPENED:

OCCURRENCES: $20 + 50 + 30$

$$P = \frac{20+50+30}{100} = P(5) + P(7) + P(9) = 1$$

IN GENERAL: $\sum_{i=1}^{M} P(x_i) = 1$

USEFUL - ONLY NEED RELATIVE PROBS:

ex $P(A) = 2P(B) = 4P(C) \Rightarrow P(A) = \frac{4}{7},\ P(B) = \frac{2}{7},\ P(C) = \frac{1}{7}$

(and A, B, C are only possible results)

(3) MULTIPLYING PROBS:

- MULTIPLE MEAS
- INDEPENDENT $\Rightarrow$ DON'T AFFECT EACH OTHER,

(COUNTER-EX: PROB OF GETTING 2 ACES IN POKER; GET 1, ONLY 3 LEFT)

- BOTH / AND

ex THROW 2 DICE, PROB OF 5 & 7 (OR 1 DIE TWICE)

EXPT: DO LOTS OF THROWS IN PARTICULAR WAY (to relate to above)

1^{ST} DIE: THROW 100 TIMES

$\Rightarrow$ GET SAME DISTR. AS ABOVE

2^{ND} DIE: FOR EACH THROW OF 1^{ST}, THROW 2^{ND} DIE 100 TIMES $\Rightarrow$ EACH SET OF 100 SHOULD LOOK ~ SAME (ELSE NOT INDEP.)

(ONLY WORKS PERFECTLY FOR ∞ # THROWS)

\# PAIRS: $N_I \times N_{II} = 100 \times 100$

\# OF $(\underset{I}{5}$ AND $\underset{II}{7})$'s : 20×50

$$P(5 \text{ \& } 7) = \frac{20 \times 50}{100 \times 100} = P_I(5) \cdot P_{II}(7) = 0.1$$

⇒ MORE RESTRICTIVE, P DECR.

IF IGNORE ORDER?

$$P(5 \text{ \& } 7) + P(7 \text{ \& } 5) = 0.2$$

NOTE: P_I & P_{II} COULD BE DIFF. DISTR FOR DIFF QTY'S

ex PROB THAT WIN LOTTERY & GET HIT BY LIGHTNING

(4) AVERAGE (OR "EXPECTATION VALUE")

$$\bar{u} = u_{AVE} \quad (\text{SOMETIMES } \langle u \rangle)$$

USUAL METHOD: ADD ALL & DIV BY N:

$$\frac{20 \times 5 + 50 \times 7 + 30 \times 9}{100}$$

$$= \underbrace{\left(\frac{20}{100}\right)}_{P(5)} \cdot 5 + \underbrace{\left(\frac{50}{100}\right)}_{P(7)} \cdot 7 + \underbrace{\left(\frac{30}{100}\right)}_{P(9)} \cdot 9$$

$$= \sum_{i=1}^{M} P(u_i) \cdot u_i$$

OTHER FNS OF u:

ex $$\overline{u^2} = \frac{20 \times (5)^2 + 50 \times (7)^2 + 30 \times (9)^2}{100} = \sum_{i=1}^{M} \underset{\text{SAME DISTR.}}{P(u_i)} \cdot u_i^2$$

GENERAL:

$$\overline{f(u)} = \sum_{i=1}^{M} P(u_i)\, f(u_i)$$

PROPERTIES (from general expression)

$$\overline{f(u)+g(u)} = \overline{f(u)} + \overline{g(u)}$$

$$\overline{c\, f(u)} = c\, \overline{f(u)}$$

(5) QTYS WHICH CHARACTERIZE P:

(a) AVE $\bar{u}$ (ALSO "MEAN")

OFTEN NEED MORE INFO:

FLUCTUATIONS: HOW MUCH DOES u JUMP AROUND?

ex 10 POLLS (treat each as measurement)

ON AVE: ~~GORE~~ KERRY 51% BUSH 49% ('04) GORE 53% BUSH 47% ('99) {put this in in Aug '99, before primaries}

McCAIN 52% CLINTON 48% ('05) OBAMA 51% McCAIN 49% ('07) ('08) {put in Spring '07, before primaries}

BET ON IT?

- HOW LIKELY TO CHG. ON NEXT POLL? (ie the one that counts)

⇒ SEE HOW MUCH 10 POLLS FLUCTUATE.

SPS. ONE POLL TO NEXT JUMPS 10%? NO.

(ASIDE: CAN DECREASE FLUCTS BY USING MORE PEOPLE IN EACH POLL; IMPORTANT LATER)

⇒ [Cf PICTURES OF DISTRIBUTIONS] all three have same $\bar{u}$, but very different distributions

(b) AVE DEVIATION:

$$\Delta u_i \equiv u_i - \bar{u}$$

$$\overline{\Delta u} = \sum_i P(u_i)(u_i - \bar{u}) \qquad \{P_i \equiv P(u_i)$$

$$= \sum_i P_i u_i - \sum P_i \bar{u}$$

$$= \bar{u} - \bar{u} \sum P_i$$

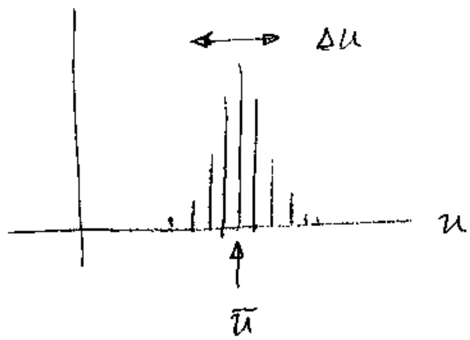
$$\boxed{\overline{\Delta u} = 0}$$

Why? as many on one side of $\bar{u}$ as other – that's why it's ave.

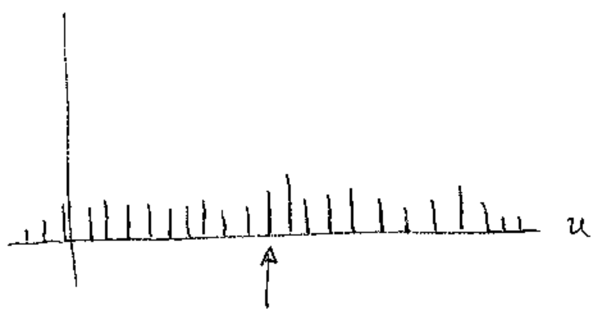
NOT USEFUL

COMPARE

$P(u)$



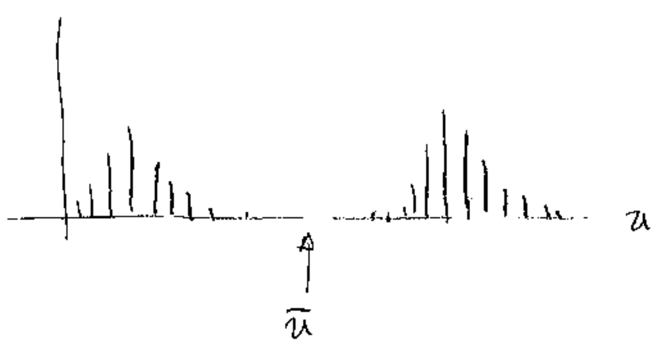
KNOW $u \sim \bar{u} \pm \Delta u$



$\bar{u}$ ALMOST MEANINGLESS

COULD GET ANYTHING

MORE PERVERSE



NEVER GET $\bar{u}$ IN ANY MEAS.

(c) VARIANCE:

(ALSO "2ND MOMENT ABOUT MEAN", OR "DISPERSION")

⇒ USE MAGNITUDES

$$\overline{(\Delta u)^2} = \sum_i P_i (\Delta u_i)^2 = \sum_i P_i (u_i - \bar{u})^2$$

USEFUL ID:

$$= \sum_i P_i \left[u_i^2 - 2\bar{u}\, u_i + \bar{u}^2 \right]$$

$$= \overline{u^2} - 2\bar{u} \underbrace{\sum P_i u_i}_{\bar{u}} + \bar{u}^2 \underbrace{\sum_i P_i}_{1}$$

$$\boxed{\underbrace{\overline{(\Delta u)^2}}_{\overline{(u-\bar{u})^2}} = \overline{u^2} - \bar{u}^2}$$

ROOT-MEAN SQUARE DEVIATION (STD. DEVIATION)

$$\sqrt{\overline{(\Delta u)^2}} \equiv \Delta^* u \quad \text{(REIF)}$$
$$\equiv \sigma \quad \textit{(statistics)}$$

very diff result for first & second P's

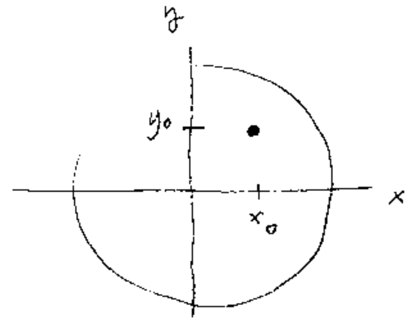
(d) HIGHER MOMENTS.

" n^{TH} MOMENT ABOUT MEAN" $\overline{(\Delta u)^n} = \sum_i P_i (\Delta u_i)^n$

- SENSITIVE TO u FURTHER FROM $\bar{u}$
- COULD TELL PLOT 2 FROM 3
- W/ ALL n, CAN RECONSTRUCT P_i

  (NOT USUALLY NEEDED)

CONTINUOUS DISTRIBUTIONS:



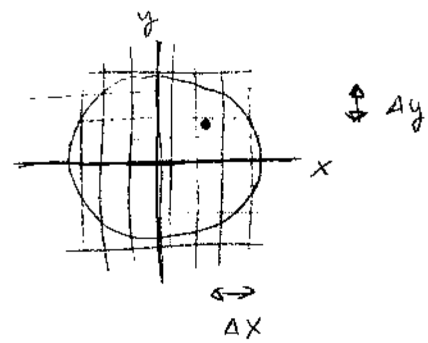
ex DARTBOARD

PROB DART HITS AT PT. $Q \equiv (x_0, y_0)$?

(darts chosen from set that hit somewhere on board)

$P(Q) = 0$ (HAVE ∞ # PTS) $\Rightarrow$ NOT MEANINGFUL

CHOICES



(A) MAKE GRID SIZE Δx, Δy

LABEL CELLS BY CENTERS Q_i

GIVE $P(Q_i)$ (ie count all w/in cell) "binning"

WELL-DEFINED, BUT P_i DEPENDS ON GRID → redo if chg. grid
(but suspect accuracy of dart thrower is indep. of grid)

(B) DEFINE PROB. DENSITY $\mathcal{P}(x, y)$

– AS DECREASE ΔX, Δy $\Rightarrow$ $P(Q_i)$ DECREASES (SMALLER AREA)

$$\mathcal{P}(x_i, y_i) = \lim_{\Delta X, \Delta y \to 0} \frac{P(Q_i)}{\underbrace{\Delta X \, \Delta y}_{\Delta A}}$$

[illegible] with μ

(prob per area)

$\Rightarrow$ dimensions?

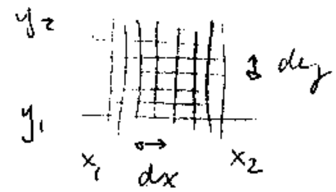
(can have well-defined limit ⇒ if half the area expect P might decrease by ~ $\frac{1}{2}$) (ie small ΔA)

don't really need → 0; it's enough to be very small vs scale at which P changes

THEN

(a) $\mathcal{P}(x, y)\, dx\, dy$ = PROB HITS FROM x to $x + dx$
y to $y + dy$

(b) PROB FROM x_1 TO x_2
y_1 TO y_2.

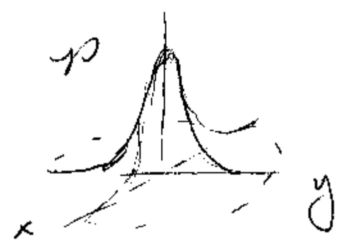


ACCEPT ANY $dx\,dy$

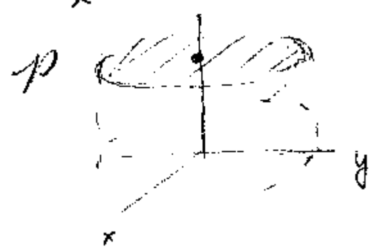
$\Rightarrow$ ADD PROBS $\mathcal{P}(x,y)\,dx\,dy$

$$\Rightarrow \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy\ \mathcal{P}(x,y)$$

ex GOOD DART THROWER:



POOR



same ave, bigger deviation

NORMALIZATION:

$$\int dx\,dy\,\mathcal{P}(x,y) = 1$$

(∞ poor dart thrower: $\mathcal{P} = \frac{1}{A} = \frac{1}{\pi R^2}$)

skip NOTE: IMPORTANCE OF SPECIFYING SET DRAW ELEMENTS FROM
(≈ constraints on your system)

(a) DARTS ON BOARD

(b) ALL DARTS (big diff. if lousy dart thrower)

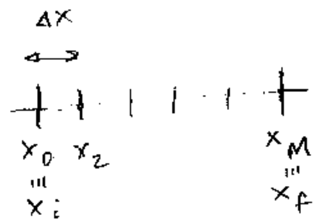
{ REIF: prob. that particular seed yields red flower → not meaningful
diff if $\in$ (a) roses
(b) all plants, flowering & not }

AVGS. (OF CONTIN. QTY'S:)

DISCRETE $\quad \bar{u} = \sum_{i=1}^{M} P(u_i)\, u_i$

CONTINUOUS (IN x)

1d



(APPROX) DISCRETE VERSION:

$$\bar{x} = \sum_{n=0}^{M-1} P(x_n)\, x_n$$

Δx → 0: $\quad \mathcal{P}(x_n) \equiv \dfrac{P(x_n)}{\Delta x} \quad$ OR $\quad P = \mathcal{P}\,\Delta x$

$$\bar{x} = \sum_{n=0}^{M-1} \Delta x\; \mathcal{P}(x_n) \cdot x_n$$

↑ skip; just find

$\mathcal{P}(x)\,dx$ = PROB. x BETW x & x+dx
mult by x & sum

$$\xrightarrow[\Delta x \to 0]{} \int_{x_i}^{x_f} \underbrace{dx\; \mathcal{P}(x)}_{\text{prob between } x \,\&\, x+dx} \cdot x \qquad \text{cf QM}$$

ALSO $\quad \overline{f(x)} = \displaystyle\int_{x_i}^{x_f} dx\; \mathcal{P}(x)\, f(x)$

2d: $\quad \overline{f(x,y)} = \displaystyle\int_{x_i}^{x_f}\int_{y_i}^{y_f} dx\,dy\; \mathcal{P}(x,y)\, f(x,y)$

RTC

PROB. DISTS WITH SEVERAL VARIABLES (skip)

SPS HAVE 2 VARIABLES u & v WHICH CAN HAVE VALUES

$$u_i \; , v_j \qquad i = 1, \cdots, M$$
$$j = 1, \cdots, N$$

CAN GIVE PROB THAT BOTH u_i & v_j OCCUR:

$$P(u_i, v_j)$$

(1) NORM:
$$\sum_{i=1}^{M} \sum_{j=1}^{N} P(u_i, v_j) = 1$$

(2) PROB THAT u_i OCCURS FOR ANY v:
$$P_u(u_i) = \sum_{j=1}^{N} P(u_i, v_j)$$
(SUM OVER ALL v'S CONSISTENT)

(NOTE $\sum_{u_i} P_u(u_i) = 1 \quad \checkmark$)

(3) $$\overline{F(u,v)} = \sum_{i,j} P(u_i, v_j)\, F(u_i, v_j)$$

(4) IF $F(u)$ INDEP OF v:

$$\overline{F(u)} = \sum_i \sum_j P(u_i, v_j) F(u_i)$$
$$= \sum_i P_u(u_i) F(u_i)$$

SPECIAL CASE: u & v STATISTICALLY <u>INDEPENDENT</u>

$\equiv$ PROB FOR u_i OCCURING DOESN'T DEP. ON v_j

THEN CAN STATE SEP. PROB FOR BOTH:

$$P(u_i, v_j) = P_u(u_i) \cdot P_v(v_j)$$ (BOTH NORMALIZED)

<u>ex</u> <u>DEPENDENT</u>: DIE WITH ODD SIDES BLUE
EVEN " RED

LET $n = 1, \cdot, 6$
$C = R, B$

COMBS

n	C
1	B
2	R
3	B etc

$P(1, R) = 0$

$P(1, B) = \frac{1}{6}$

$P(n < 4, B) = \frac{2}{6} = \frac{1}{3}$

<u>INDEP</u>: 1 DIE, & 1 COIN (B & R)

COMBS

1	B
1	R
2	B
2	R

etc

$P(1, R) = P(1) \cdot P(R) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$

$P(1) \cdot P(B) = \frac{1}{12}$

$P(n < 4, B) = \frac{3}{6} \cdot \frac{1}{2} = \frac{1}{4}$