starting in lecture notes part (2) | STAT 206 14 Jan 21

p. (18) middle | trading off false positives against false negatives in disease screening (FP) (FN) | lecture (1)

p. (19) bottom should be

FP = (test says ⊕ AND really ⊖)

FN = (test says ⊖ AND really ⊕)

(B.D) p. (20) middle

(B.D) p. (18) truth

$u(a, \theta)$

		$\theta = 1$	$\theta = 0$
test says	⊕	0	$-L$
	⊖	$-c \cdot L$	0

(c >> 1)

0	-1
$-c$	0

ok to divide by L because utility is ratio scale (0 = no loss):

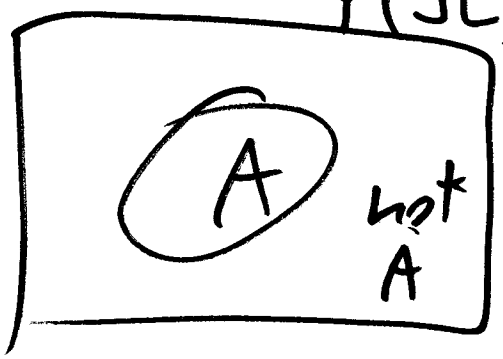
conjecture: there's a sense in which screening test designers have used c = 594 (I'll check this later)

999	594
1	98,406

$P(A|B) = ?$

$P(A|B) + P(\text{not}A|B) = 1$

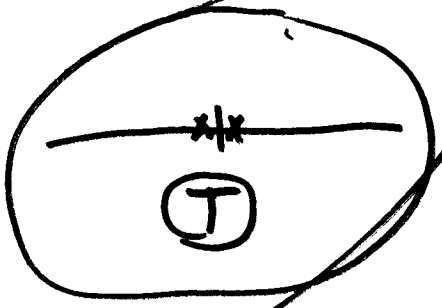
$P(\Omega) = 1$



$P(A|B) =$

$1 - P(\text{not}A|B)$

$P(\theta = 1 | y_1 = 1, B) \doteq 0.63$



$P(y_1 = 1 | \theta = 1, B) = .999$

"truth" = gold standard = Western Blot

(quite a bit more accurate than "Determine" test but a lot more expensive [time money])

~~cost high, accuracy~~

a_1 = (only "Det.") cost low, accuracy low a_2 = (only w. Blot) high

"Determine" almost always right when it says ⊖

a_3 = (run Det. first; if ⊖, stop & tell Bob ⊖; if ⊕, run wBlot; if wBlot says ⊕, tell Bob ⊕; if wB⊖, tell Bob ⊖)

adaptive

looking for optimal tradeoff of costs ($) against benefits (accuracy)

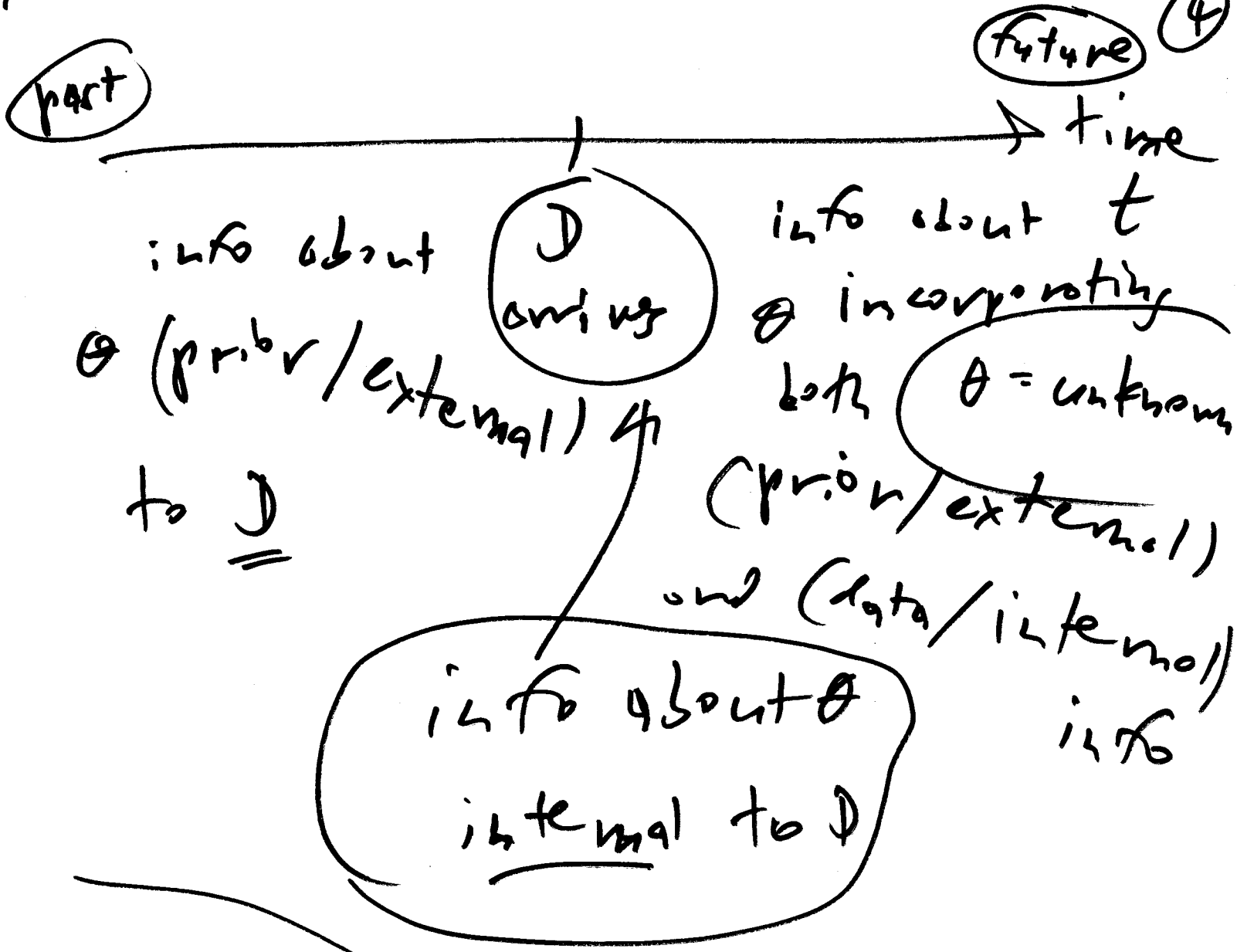
how convert accuracy to $?

$p \xrightarrow{\text{unique}} (\mathbb{R}, \mathbb{C}) \xrightarrow{\text{not necessarily unique}} (\Theta, \mathcal{D}, \mathcal{B})$

model

uncertain (how specify?)

$M_{Ip} = \{ p(\Theta|\mathcal{B}), p(\mathcal{D}|\Theta, \mathcal{B}) \}$ uncertain



7 activities in statistical data science

- probability calculations
  pretend θ known, work out $p(D|\theta_B)$
- statistical activities:

5 things: { 4 on p. 35 (LN part 3)

① data curation ② ~~③~~ design of data-gathering