This time: Bayes

next time: random variables

false positive rate $= P(\text{not } A \mid \oplus) = 1 - P(A \mid \oplus) \doteq 89\%$

AMS 131 7 Aug 19

1

$$\frac{P(A \mid \oplus)}{P(\text{not } A \mid \oplus)} = \left[\frac{\overset{.004}{P(A)}}{\underset{.996}{P(\text{not } A)}}\right]\left[\frac{\overset{0.96}{P(\oplus \mid A)}}{P(\oplus \mid \text{not } A)}\right]$$

1 − specificity 0.03

$P(A) = p$

$o_A = \frac{p_A}{1 - p_A}$

odds ratio in favor of A

$$= \frac{(.004)(.96)}{(.996)(.03)}$$

$$= \left(\begin{array}{c} 249 \text{ to } 1 \\ \text{against} \\ \text{really} \\ \text{HIV}\oplus \end{array}\right)\left(\begin{array}{c} \text{data} \\ \frac{96}{3} \text{ odds} \\ \text{in favor} \\ \text{of really HIV}\oplus \\ 32 \text{ to } 1 \\ \text{in favor} \end{array}\right)$$

$P(\text{not } A) = 1 - p_A$

$o_{\text{not } A}$ = odds in favor of not A $= \frac{1}{o_A}$

$$\left(\text{posterior odds ratio in favor of } A\right) = \frac{\cancel{249} \cdot 96}{3 \cdot 249} = \frac{96}{747}$$

$$= o_A$$

$$p_A = \frac{o_A}{1 + o_A}$$

method 3: evaluate

$$11\% \doteq \frac{96}{843} = \frac{\frac{96}{747}}{1 + \frac{96}{747}}$$

$P(+ \mid \text{data})$ hard but $P(+ \mid A)$ $P(+ \mid \text{not } A)$ easy

A, not A	truth
$+$, $-$	data

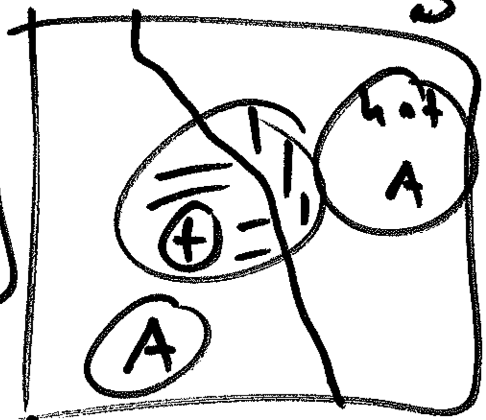
D.V. Lindley: "extending the conversation"

to compute probabilities about data, bring in truth partition over

S ③

$P(\oplus) =$

$P[(\oplus \text{ and } A) \text{ or } (\oplus \text{ and not } A)]$



$= P(\oplus \text{ and } A) + P(\oplus \text{ and not } A)$

(A, not A) = a partition

$= P(A)\,P(\oplus|A) + P(\text{not } A)\,P(\oplus|\text{not } A)$

$= (.004)(.96) + (.996)(.03) = .03372$

Bayes' theorem, again

partition sets $(C_1, \ldots, C_k)$

$$P(B|A) = \frac{P(B)\,P(A|B)}{\sum_{i=1}^{k} P(C_i)\,P(A|C_i)}$$

LTP

more general than necessary for us for now for now

ELISA

B_1 = really (A) is HIV ⊕

B_2 = really is HIV ⊖ (not A)

truth is one of $B_1, \dots, B_k$ (unknown states of world), and $(B_1, \dots, B_k)$ forms a partition; then

$$P(\underbrace{B_i}_{\text{truth}} \mid \underbrace{A}_{\text{data}}) = \frac{P(B_i)\,P(A \mid B_i)}{\sum_{j=1}^{k} P(B_j)\,P(A \mid B_j)} \qquad (9.50)$$

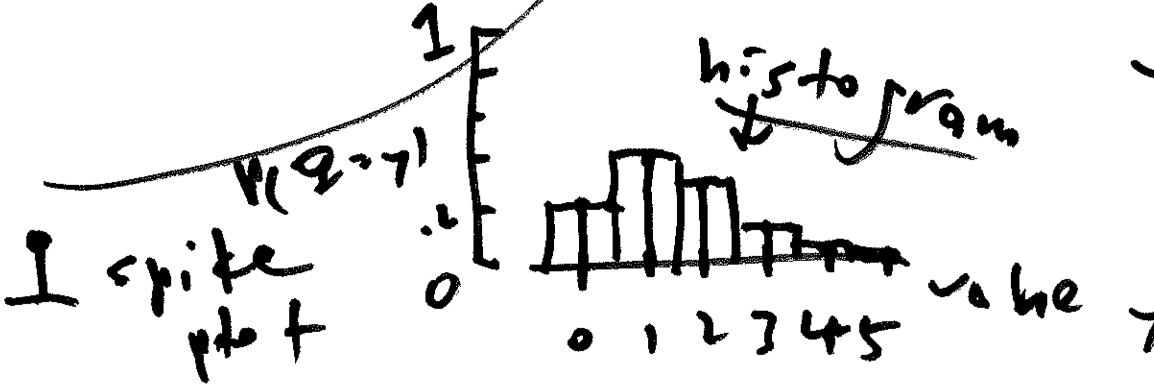
(H, Chimp)

H   chimp

$P(Y = y)$

Y: r.v.

y: a possible value of Y



Y has the Bernoulli(p) dist. if

PMF of Y

$$f_Y(y) = p^y (1-p)^{1-y} I_{\{0,1\}}(y)$$

Support of Y

T/F

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{false} \end{cases}$$

$$I_A(y) = \begin{cases} 1 & \text{if } y \in A \\ 0 & \text{else} \end{cases}$$

set

"is distributed as"

$Y \sim$ Bernoulli(p) (Freq.)

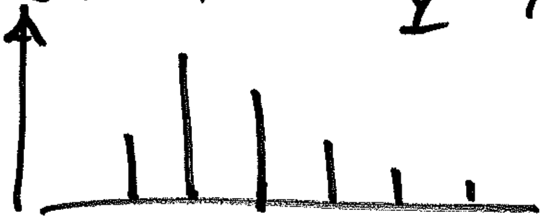
Y has the Bernoulli(p) dist.

$(Y \mid p) \sim$ Bernoulli(p) (Bayesian)

(10.53)

discrete RV have PMFs

$P(Y=y)=f_Y(y)$ ← probability mass functions

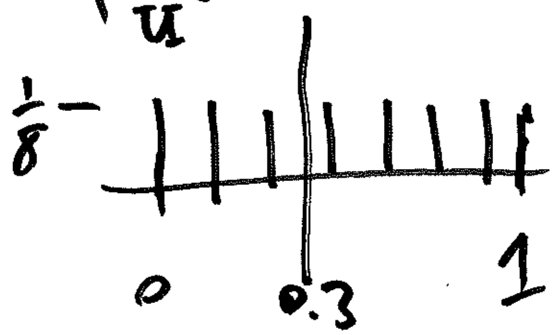


y (possible values)

PMF

$f_Y(y) \ge 0$ and $\sum_{\text{all } y} f_Y(y) = 1$

$f_U(u)$



PMF

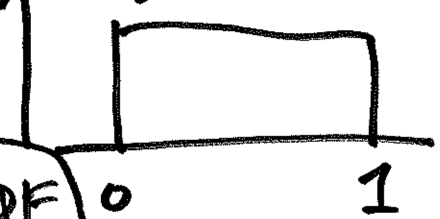
$P(U \le 0.3) =$

$P(U = \frac{0}{8} \text{ or } \frac{1}{8} \text{ or } \frac{2}{8})$

$\sum_{u=\frac{0}{8},\frac{1}{8},\frac{2}{8}} f_U(u)$

continuous RV have PDFs (probability density function)s

$f_U(u)$ ←

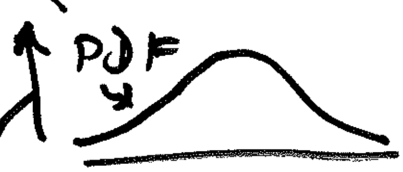


PDF

and

$f_Y(y) \ge 0$ $\int_{-\infty}^{\infty} f_Y(y) = 1$ $f_Y(y)$

or



bell curve

Normal (Gaussian) dist.

$$P(a \le Y \le b) = \int_a^b f_Y(y)\,dy$$

← cont. f'n

$$\int_a^a f_Y(y)\,dy = 0$$

if Y is a cont. RV

$$\therefore \quad P(Y = a) = 0 \quad \text{for all } a \in \mathbb{R}$$

therefore

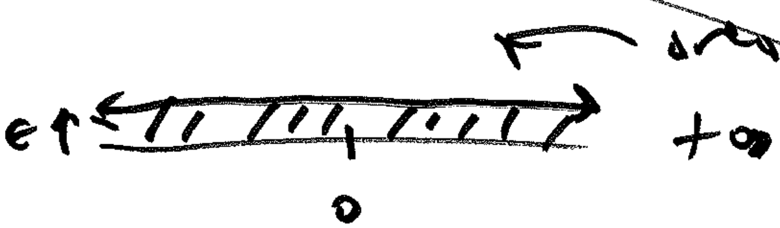
singleton $\{a\}$

$\epsilon > 0$

$$\int_{-\infty}^{\infty} \epsilon\,dy = \epsilon y \Big|_{-\infty}^{\infty}$$

$$= \epsilon\,[\infty - (-\infty)]$$

$$= \infty$$



$$P(a \le Y \le b) = \int_a^b f_Y(y)\,dy$$

prob. for cont. RV $\leftrightarrow$ area under pdf