BASIC DISTRIBUTIONS, ENTROPY AND PROBABILITIES

Ensembles

X ensemble is a triple system:

$$(\underbrace{x}_{\text{random variable}}, \underbrace{A_x}_{\text{possible value set}}, \overbrace{P_x}^{\text{probabilities}}) \qquad P_x = \{p_1, p_2, p_3, \ldots, p_I\}$$

$$A_x = \{a_1, a_2, \ldots, a_I\}$$

$$\underbrace{P(x=a_i)}_{\substack{p(a_i)\\ p(x)}} = p_i, \quad p_i \geq 0, \quad \sum_{a_i \in A_x} P(x=a_i) = 1$$

Example: Frequency of letters of the alphabet

Sub-ensemble probabilities

If T is a sub ensemble of A_x, then:

$$P(T) = P(x \in T) = \sum_{a_i \in T} P(x=a_i)$$

(the probabilities are added / summed)

Joint Ensemble

$\underset{\text{(ordered)}}{x, y}$; $x \in A_x = \{a_1, \ldots, a_I\}$, $y \in A_y = \{b_1, \ldots, b_J\}$

$P(x,y)$: The joint probability of x and y

(Notation: $x, y \leftrightarrow xy$)

Marginal Probability

$$P(x=a_i) \equiv \sum_{y \in A_y} P(x=a_i, y)$$

$$P(y) \equiv \sum_{x \in A_x} P(x,y)$$

Conditional Probability

$$P(x=a_i \mid y=b_J) = \frac{P(x=a_i, y=b_J)}{P(y=b_J)}$$

: "Given that $y=b_J$, what is the probability that $x=a_i$?"

Example: Letters in an alphabet two form xy (a word with two letters)

$\underbrace{P(xy)}_{\text{Joint (Fig. 2-2)}} : \underbrace{P(x) = P(y)}_{\text{marginal probabilities} \rightarrow \text{Fig 2.1}}$ | Joint Ensemble

- Given $x=q$, the most probable y?

  $P(y \mid x=q) = ?$ ("u", "□") ← space → Conditional (Fig 2-3)

- Given $y=u$, the most probable x?

  $P(x \mid y=u) = ?$ ("n", "o")

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928

Fig. 2-1 Probability that a random character (a–z & space) will be (Mackay) $P(x)$

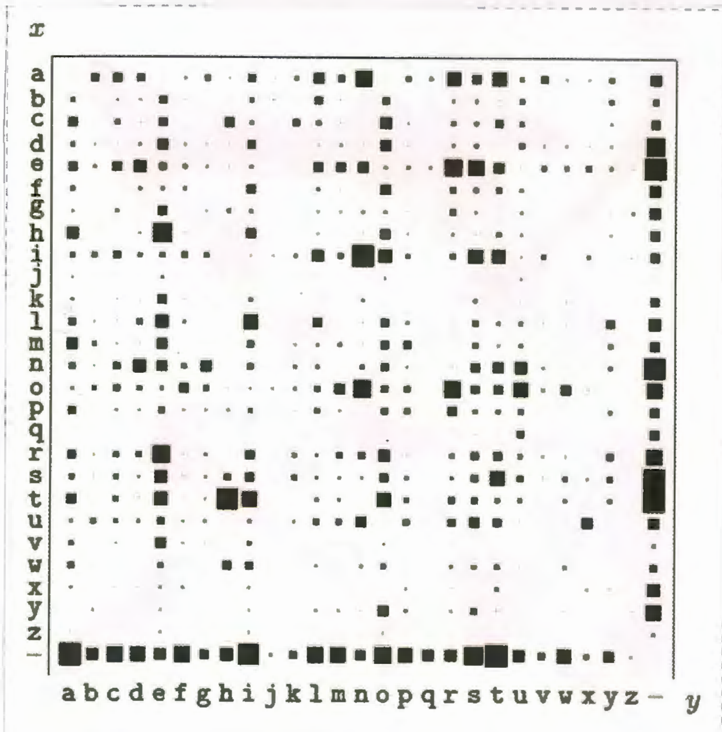
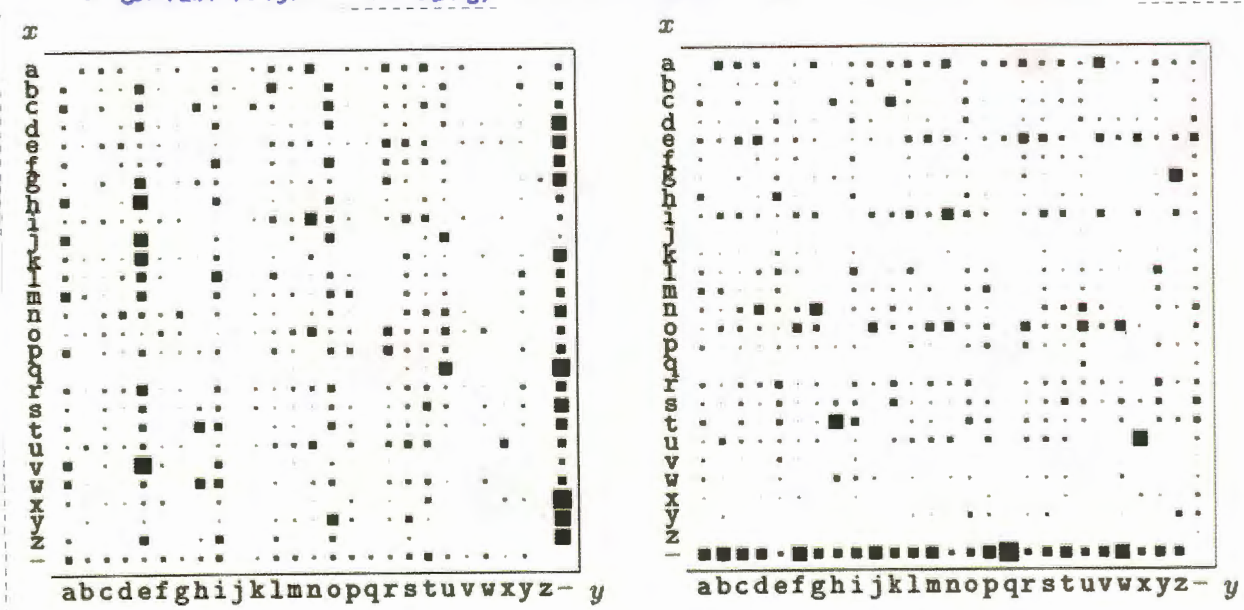


Fig: 2-2: Prob. dist. over the 27x27 possible bigrams xy (Mackay)
Joint dist $P(xy)$



(a) $P(y \mid x)$ (b) $P(x \mid y)$

Fig 2-3: Conditional Prob. dist. Each column shows the conditional distribution of the first letter, given the second letter (Mackay)

H: Universe (given everything else as it is)

Multiplication Rule: $P(x,y|H) = P(x|y,H)\,P(y|H) = P(y|x,H)\,P(x|H)$

Summation Rule:
$$P(x|H) = \sum_y P(x,y|H)$$
$$= \sum_y P(x|y,H)\,P(y|H)$$

Bayesian Rule:
$$P(y|x,H) = \frac{P(x|y,H)\,P(y,H)}{P(x|H)}$$
$$= \frac{P(x|y,H)\,P(y|H)}{\sum_{y'} P(x|y',H)\,P(y'|H)}$$

Independency: For x and Y to be considered independent of each other:
$$X \perp Y \longleftrightarrow P(x,y) = P(x)\,P(y)$$

<u>Example</u> Jo takes a medical exam.

Variables are:

a: Jo's health
- a = 1 : Jo is sick
- a = Ø : Jo is healthy

b: Test's result
- b = 1 : Test result is true
- b = Ø : Test Result is false

<u>Data:</u>
* Test is 95% reliable
* The probability that somebody at Jo's age, status and backgrand to be sick is 1%

Jo takes the test and the result turns out to be positive (indicating that Jo is sick). What is the probability that Jo is indeed sick?

Conditional
$$P(b=1|a=1) = 0.95 \qquad P(b=1|a=0) = 0.05$$
$$P(b=0|a=1) = 0.05 \qquad P(b=0|a=0) = 0.95$$

Marginal
$$P(a=1) = 0.01, \quad P(a=0) = 0.99$$

(Example cont'd...)

$$P(a,b) = P(a)P(b|a), \quad P(b=1) = P(b=1|a=1)P(a=1) + P(b=1|a=0)P(a=0)$$

$$P(a=1|b=1) = \frac{P(b=1|a=1)P(a=1)}{P(b=1|a=1)P(a=1) + P(b=1|a=0)P(a=0)}$$

$$= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99}$$

$= 0.16 \longrightarrow$ 16% probability that Jo is indeed sick.

Information Content

Suppose that we want to use a coding similar to that of the Morse Coding. Should we assign a single dot to "a" or "z"?

We must assign the shortest symbol (like "." or "–") to the most frequently used letters and long symbols (like "..." or ".–.") to the infrequently used letters to reduce transmission time. This is reflected and calculated by the information content.

$h(x) = \log_2 \frac{1}{P(x)}$ it's measured in units of bits (0 or 1)

Entropy

Entropy, in this context, is defined as the average information content:

$$S(x) = \sum_{x \in A_x} P(x) \log \frac{1}{P(x)}$$

$\longrightarrow S(x) \geq 0$ (only equal to 0 when $p_i = 1$)

$\longrightarrow$ entropy is maximum when P is uniform (equal probabilities)

Expected (Average) Value: $\mu = \sum_{x_i} x_i \cdot p(x_i) \qquad (\langle x \rangle, E[x])$

$$\{x\} = \{1, 3, 3, 5, 7, 7, 7, 8, 9, 9\}$$

$$\langle x \rangle = \frac{1+3+3+5+7+7+7+8+9+9}{10} = 5.9 \quad \text{(conventional method)}$$

$$\left.\begin{array}{ll} P(x_i=1) = \frac{1}{10} = 0.1 & P(x_i=7) = 0.3 \\ P(x_i=2) = 0 & P(x_i=8) = 0.1 \\ P(x_i=3) = \frac{2}{10} = 0.2 & P(x_i=9) = 0.2 \\ P(x_i=5) = 0.1 & \end{array}\right\} \quad \langle x \rangle = \sum_{x_i} x_i\, p(x_i)$$

$$= 0{,}1 \cdot 1 + 0{,}2 \cdot 3 + 0{,}1 \cdot 5 + 0{,}3 \cdot 7 + 0{,}1 \cdot 8 + 0{,}2 \cdot 9 = 5.9$$

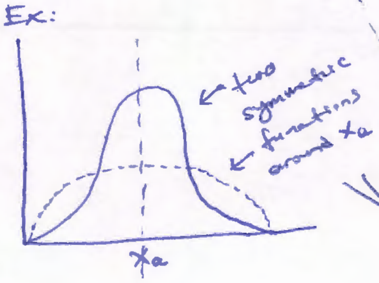
(systematic)

Variance: $\sigma^2 = \sum_{x_i} x_i^2\, p(x_i) - \mu^2$

Standard Deviation: $\sigma = \sqrt{\sigma^2}$

$$\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \sqrt{\sum_i p_i (x_i - \mu)^2}$$

Ex:



$\mu_1 = \mu_2 = x_a$

$\sigma_1 \neq \sigma_2$