

Last time

- Expectation

Today

- Applications of Expectation
 - Variance & Standard Deviation
 - Entropy

Next time

- Conditional Prob.
- Bayes Theorem

Variance

Consider heights

case 1

4'10" 5' 5'2"

$$E\{x\} = 5'$$

case 2

4' 5' 6'

$$E\{x\} = 5'$$

case 3

3' 5' 7'

$$E\{x\} = 5'$$

How to measure "variability" 3 ways

case 2

$$\textcircled{1} E\{Y\} = \sum_{w \in \Omega} Y(w) \cdot p(w)$$

$$= (-12'') \cdot \frac{1}{3} + (0'') \cdot \frac{1}{3} + (+12'') \cdot \frac{1}{3}$$
$$= 0$$

$$\textcircled{2} E\{Y\} = |-12''| \cdot \frac{1}{3} + |0''| \cdot \frac{1}{3} + |12''| \cdot \frac{1}{3}$$
$$= 12 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 12 \cdot \frac{1}{3} = 8 \text{ inches}$$

$$\textcircled{3} E\{Y\} = E\{(X - E\{X\})^2\} = \text{case 1}$$
$$= (-2'')^2 \cdot \frac{1}{3} + (0'')^2 \cdot \frac{1}{3} + (+2'')^2 \cdot \frac{1}{3}$$
$$= \frac{4}{3} + 0 + \frac{4}{3} = \frac{8}{3} \text{ inches}^2$$

X neg. & pos. cancel

- mean absolute deviation

- variance

variance σ^2

Take square root, get back inches...

$$\sigma^2 = \text{Var}(X) = E[(X - E\{X\})^2]$$

$$\Rightarrow \sigma = \sqrt{\text{Var}(X)} = \sqrt{E[(X - E\{X\})^2]} \quad - \text{ back in units originally measured.}$$

e.g. case 1: $\sigma^2 = \text{Var}(X) = 8/3 \text{ inches}^2$

$$\sigma = \sqrt{\text{Var}(X)} = \sqrt{8/3 \text{ inches}^2} = 1.63''$$

case 3: ... $\sigma = 19.6''$

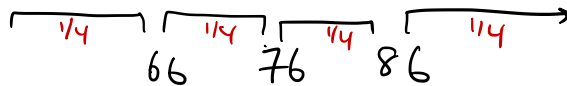
Exam

$$E\{X\} = 76$$

$$\sigma = 15$$

$\pm 2/3 \sigma$ - quantiles

$$2/3 \sigma = 10 \text{ pts}$$



Entropy

Consider 8 letters only $\{A, B, C, D, E, F, G, H\}$

Need 3-bits

A $\rightarrow$ 000

B $\rightarrow$ 001

H $\rightarrow$ 111

} fixed length code

suppose must encode n -things, need: $\lceil \log_2 n \rceil$

Efficiency of code is measured in bits-per-character on average, BPC. $BPC = 3$

variable length code: assign short codes to frequent letters
long codes to infrequent letters.

$$BPC = \sum_{w \in \mathcal{L}} X(w) \cdot P(w) = \sum_i l_i \cdot p_i$$

Why must we have

long codes to make up for short codes?

Analog to PHP: Kraft's Inequality

Example:

A B C D E F G H

BBB $\rightarrow$ 010101

00 01 010 011 100 101 110 111

CF $\rightarrow$ 010101

code length for i th letter

probability of i th letter

Suppose we have following skewed distribution

$$P_i = \left(\overset{A}{1/4}, \overset{B}{1/4}, \overset{C}{1/8}, \overset{D}{1/8}, \overset{E}{1/16}, \overset{F}{1/16}, \overset{G}{1/16}, \overset{H}{1/16} \right)$$

$$(00, 01, 100, 101, 1100, 1101, 1110, 1111)$$

$$ADF \rightarrow 001011101$$

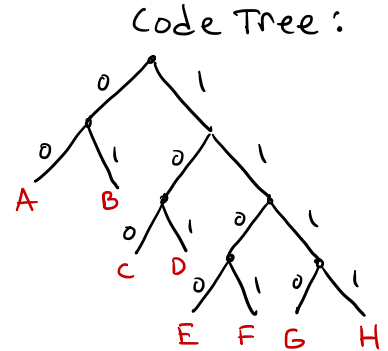
$$\begin{array}{c} 00|10|11101 \\ A \quad D \quad F \end{array}$$

$$BPC = \sum_i l_i \cdot p_i$$

$$= 2 \cdot 1/4 + 2 \cdot 1/4 + 3 \cdot 1/8 + 3 \cdot 1/8 + 4 \cdot 1/16 + \dots + 4 \cdot 1/16$$

$$= 2.75$$

* 8.33% savings over fixed code of length 3.



$$P_2 = \begin{matrix} A & B & C & D & E & F & G & H \\ \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64} \right) \end{matrix}$$

0 10 110 1110 111100 111101 111110 111111

$$BPC = \sum_i l_i \cdot p_i$$

$$= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 6 \cdot \frac{1}{64} + \dots + 6 \cdot \frac{1}{64}$$

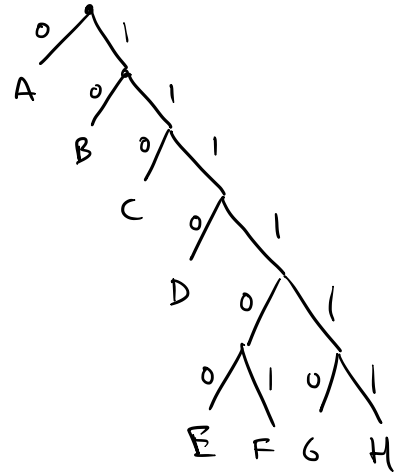
$$= 2$$

p_i	l_i
$\frac{1}{2}$	1
$\frac{1}{4}$	2
$\frac{1}{8}$	3
$\frac{1}{16}$	4
$\frac{1}{64}$	6

$$p_i = \frac{1}{2^k} \Rightarrow p_i = \frac{1}{2^{l_i}}$$

$$\Leftrightarrow 2^{l_i} = \frac{1}{p_i}$$

$$\Leftrightarrow l_i = \log_2 \frac{1}{p_i}$$



Entropy.

$$Q: \quad P = (p_1, p_2, p_3, \dots)$$

$$l_i = \log_2 \frac{1}{p_i}$$

What is expected code length?

$$BPC = \sum_i p_i \cdot l_i$$

$$= \sum_i p_i \cdot \log_2 \frac{1}{p_i}$$

Entropy

$$H(P) = \sum_i p_i \cdot \log_2 \frac{1}{p_i}$$

ACBADAFGAH

$$P_A = \frac{4}{10}$$

$$P_B = \frac{1}{10}$$

$$P_C = \frac{1}{10}$$

$$P_D = \frac{1}{10}$$

⋮

Facts

- ① Entropy is the compression limit.
- ② Entropy is essential to bound an ability to communicate in presence of noise.
- ③ Fundamental measure of randomness.
- ④ Cryptography

Conditional Probability

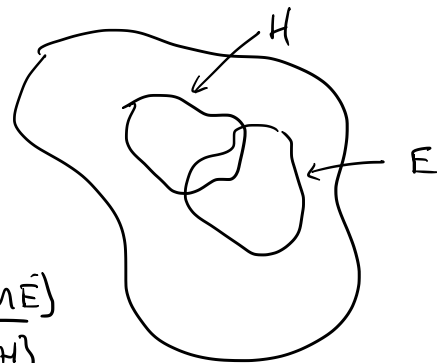
← prob of E
given H

$$Pr[H \cap E] = Pr[H] \cdot \underline{Pr[E|H]} \quad \leftarrow$$

$$\Rightarrow Pr[E|H] = \frac{Pr[H \cap E]}{Pr[H]}$$

$$= Pr[E] \cdot \underline{Pr[H|E]} \quad \leftarrow$$

$$\Rightarrow Pr[H|E] = \frac{Pr[H \cap E]}{Pr[E]}$$



~~~~~  
Bayes Law:

$$\underline{Pr[E]} \cdot \underline{Pr[H|E]} = Pr[H \cap E] = \underline{Pr[H]} \cdot \underline{Pr[E|H]}$$

$$Pr[H|E] = \frac{Pr[E|H] \cdot Pr[H]}{Pr[E]}$$

H = hypothesis: do have Zika?

E = evidence: did blood test come back pos?