Probability and Statistics

Module 1: Topic 4
Probability: Information content in data

Konduri Aditya
Department of Computational and Data Sciences
Indian Institute of Science, Bengaluru

CDS
Department of Computational and Data Sciences

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Outline

- Information of an event
- Information entropy ✓
- Mutual information ✓

Information theory

- It is a field of study concerned with quantifying information for communication
- A foundational concept from information is the quantification of the amount of information in things like events, random variables, and distributions
- Use the concepts from probability

Konduri Aditya, CDS, IISc Bengaluru

Information theory

- The sky colour is blue ✓
- The sky colour is red

Second statement provides more info.
(and indicates that something is abnormal)

- Person A always goes for jogging at 6 PM ✓
- Person B goes after the sunset ✓

This provides more information
– Season. (predict) !

Information of an event

How likely?
→ Define this in terms of a probability law.

- Quantifying information is the idea of measuring how much surprise there is in an event
- *"The basic intuition behind information theory is that learning that an unlikely event has occurred is more informative than learning that a likely event has occurred."*
- Low probability (rare) events: more surprising: **high** information
- High probability (common) events: less surprising: **low** information
- Shannon information or information h(x):

$$h(x) = -\log_2(p(x)) \quad bits$$

$[0, \infty)$ non-negative. $[0, 1]$

If $P(x) = 0$, $h(x) = \infty$

$P(x) = 1$, $h(x) = 0$

$\log_e$ / $\log_{10}$

Konduri Aditya, CDS, IISc Bengaluru

Information of an event

0, 1

Example 1: If p(x)=1 Certain event.

$$h(x) = -\,log_2(1) = 0 \quad bits$$

Example 2: coin toss (Fair)

Event (E): obtain a head

P(E) = 0.5 and $h(E) = -\,log_2(0.5) = 1.0 \quad bits$ → binary digits.

Example 3: Die roll (6-faced die, 6 outcomes)

Event (E): obtain a 6

P(E) = 1/6 and $h(E) = -\,log_2(1/6) = 2.585 \quad bits$

Die roll:

- possible outcomes

1 2 3 | $\boxed{4 \; \boxed{5} \; \boxed{6}}$

→ Naive way – 6 Qs.

1 Q.

- Divide and conquer.

(1) ≤ 3 ?

(2) ≤ 5 ?

} → 2 Q.

Information for an event

→ Outcome 6

$P(6) = 1/6$

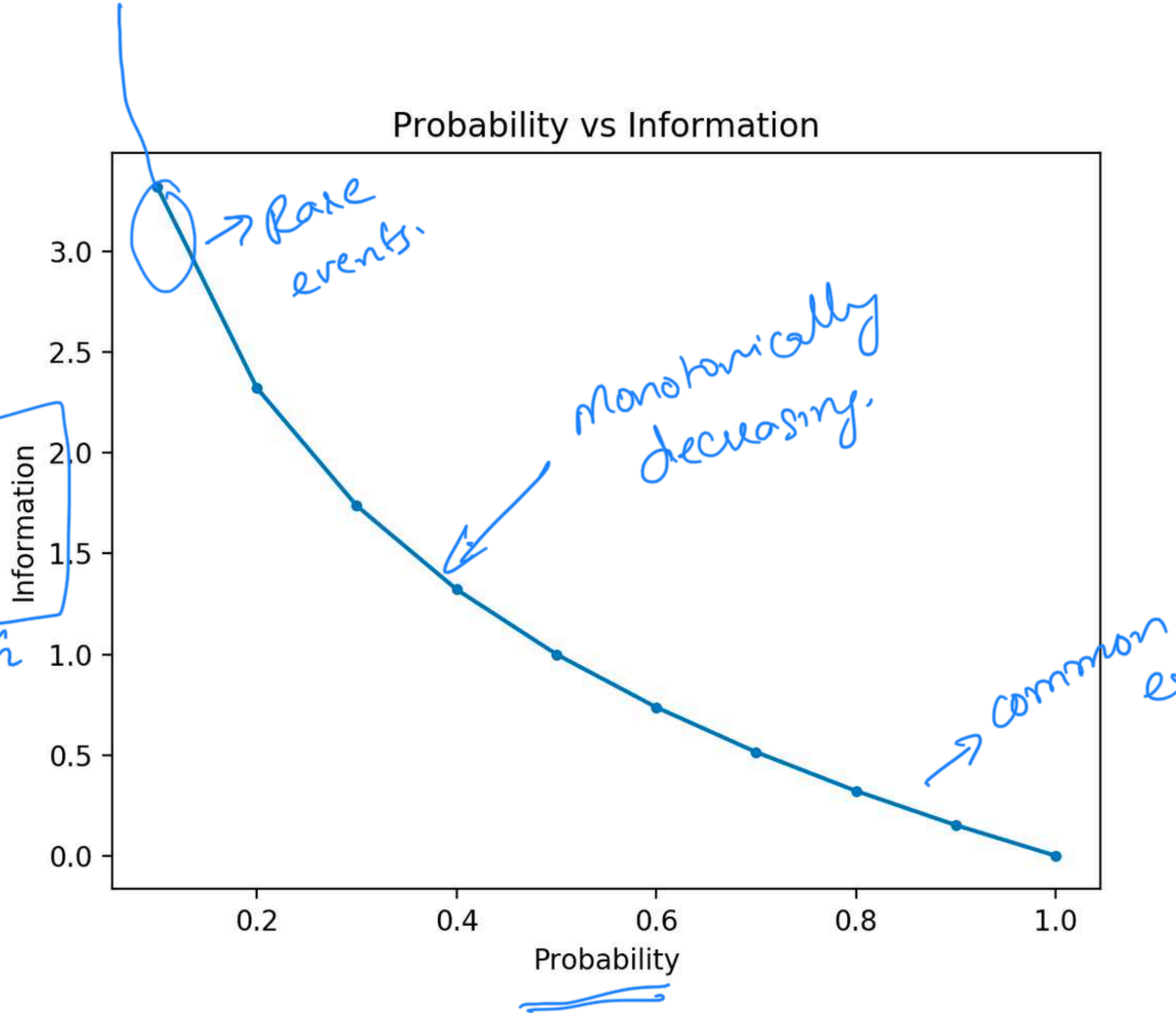
$h(x) = 2.585$

→ Outcome Even

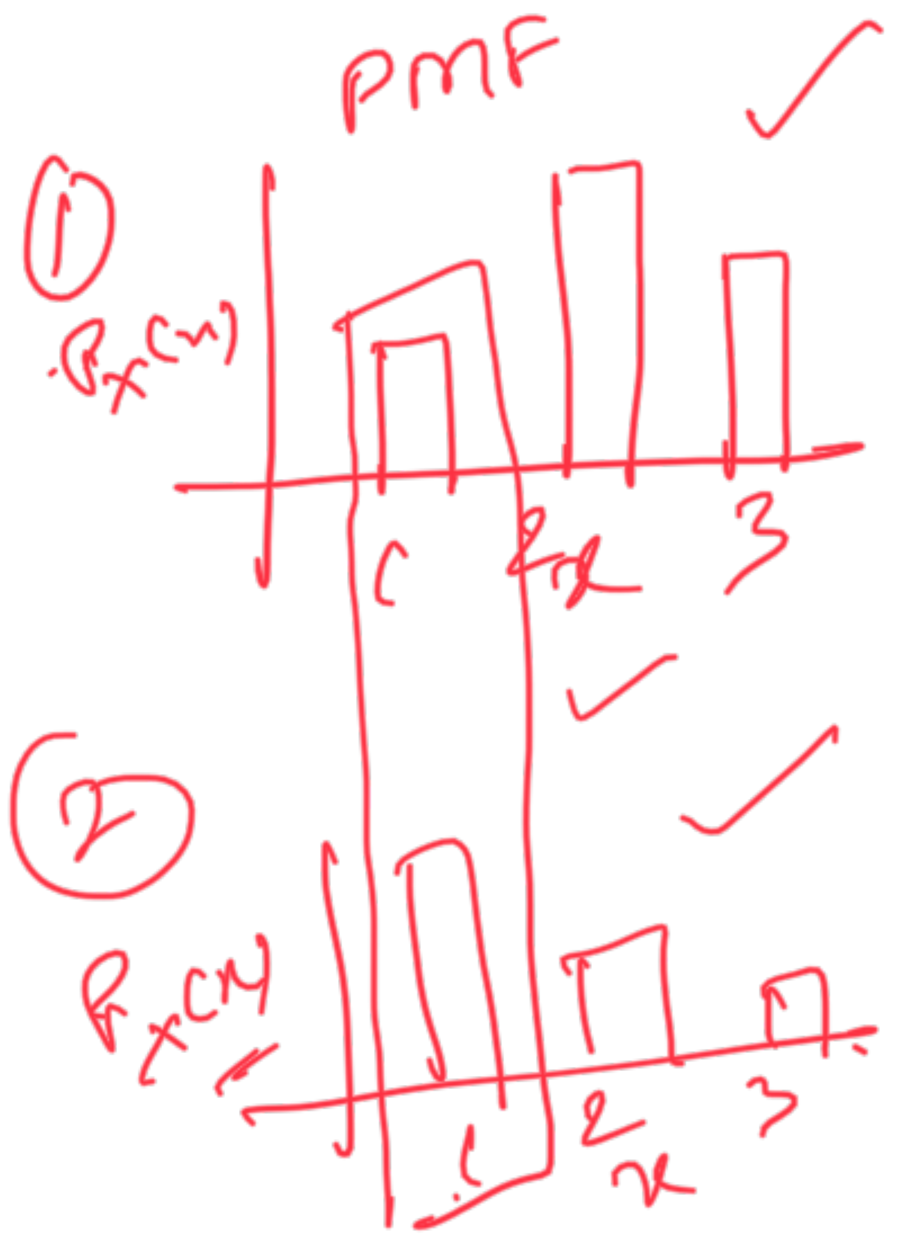
$P(\text{Even}) = \frac{1}{2}$

$h(\text{Even}) = -\log_2 \frac{1}{2}$

$= 1$



https://machinelearningmastery.com/what-is-information-entropy/



Information entropy

- Next we want to quantify how much information there is in a **random variable**
- Consider a random variable **X** with a probability distribution **P**
- Calculating information for a R.V. is called **information entropy or Shannon entropy** $\boxed{H(X)}$
- **Intuition:** it is the average number of bits required to represent or transmit an event drawn from the probability distribution for the R.V.
- If there are K discrete states (Cointoss → two discrete states)

$$H(X) = -\sum_{k=1}^{K} p(k) log_2(p(k)) \quad bits$$

Information entropy

Die roll

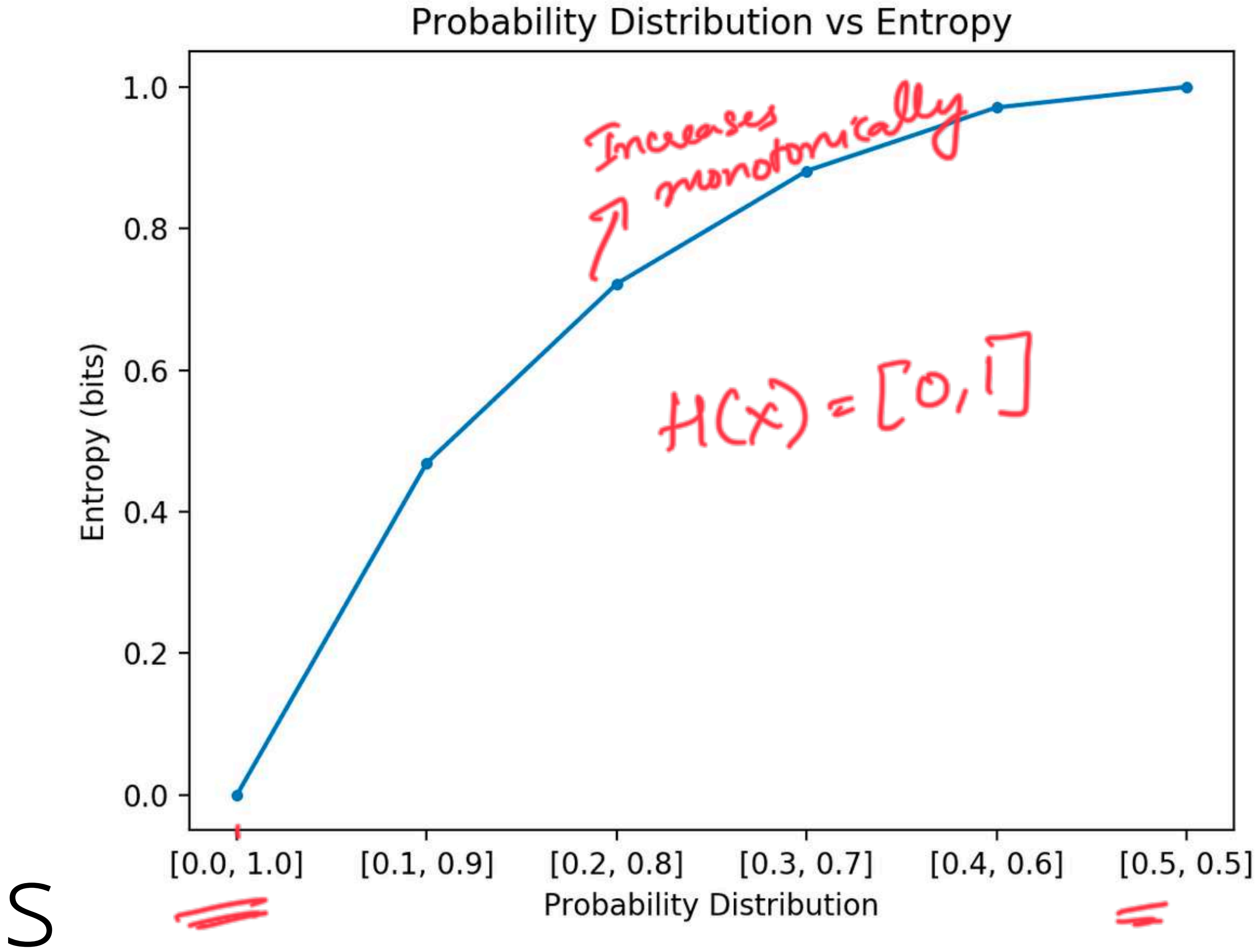
- Lowest entropy: R.V. that has a single event with a probability of 1.0
- Highest entropy: If all events are equally likely
- Example: Coin toss

$P(x=0) = 0.5$

$P(x=1) = 0.5$

  - Fair coin: X={0,1} —> no of heads

$$H(X) = -[0.5log_2(0.5) + 0.5log_2(0.5)] = 1.0 \quad bits$$



$P(x=1) = 0.75$ $P(x=0) = 0.25$

  - Biased coin (0.75,0.25): X={0,1} —> no of heads

H T

$$H(X) = -[\underbrace{0.25log_2(0.25)}_{X=0} + \underbrace{0.75log_2(0.75)}_{X=1}] = 0.8115 \quad bits$$

Skewed probability distribution: low entropy

Balanced probability distribution: high entropy

Information entropy

- Consider two random variables X and Y with a PMF $P_{X,Y}(x, y)$
- Joint entropy:

$$H(X, Y) = -\sum_{x,y} P_{X,Y}(x, y) log_2(P_{X,Y}(x, y)) \quad bits$$

Conditional entropy:

$$H(X|Y) = -\sum_{x,y} P_{X,Y}(x, y) log_2(P_{X,Y}(x|y)) \quad bits$$

The conditional entropy is a measure of how much uncertainty remains about the random variable X when we know the value of Y

Mutual information

- Entropy provides the basis for calculating the difference between two probability distributions
- Mutual information is calculated between two variables and measures the reduction in uncertainty for one variable given a known value of the other variable → Condition Y
- The mutual information between two random variables *X* and *Y*:

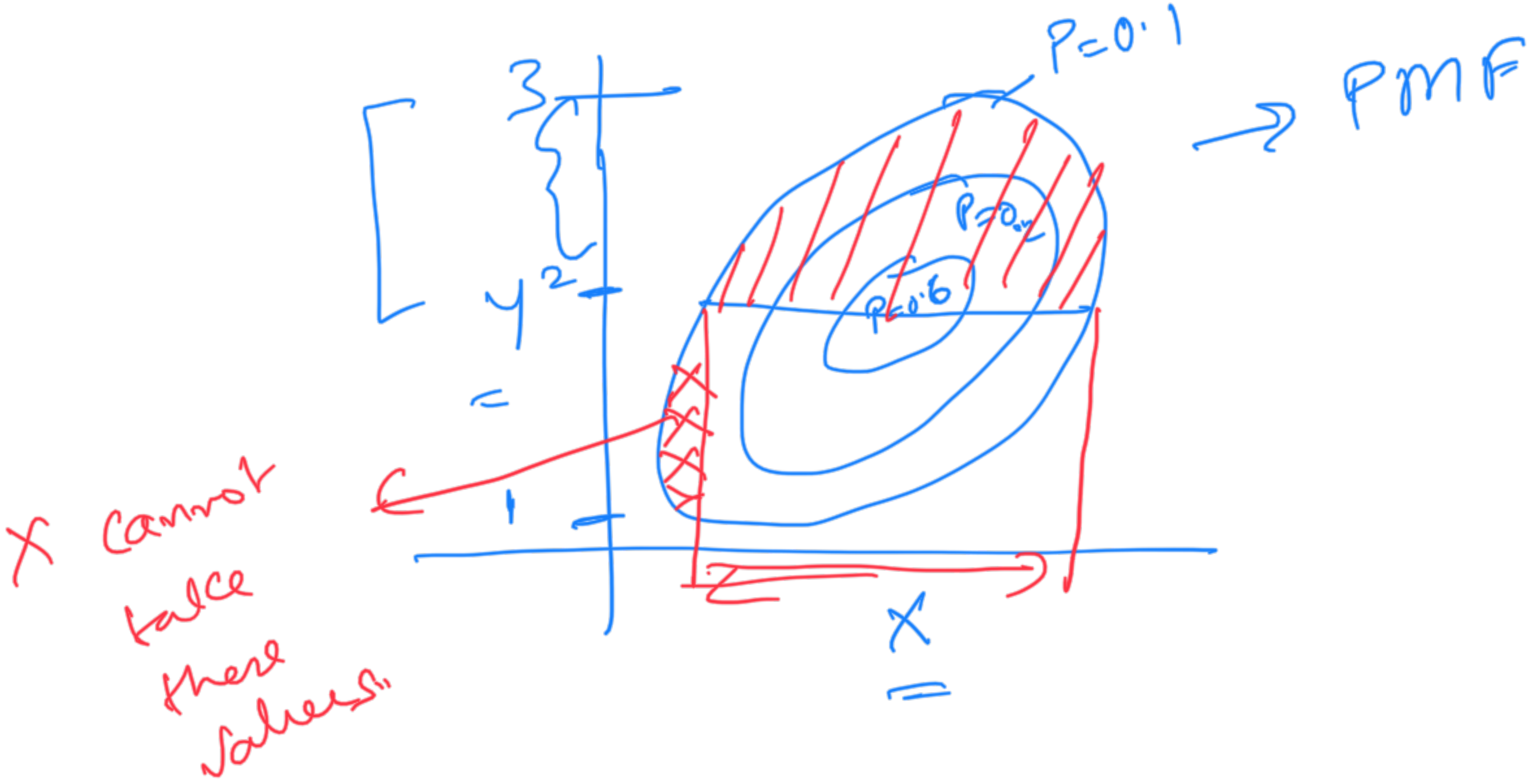
R.V.

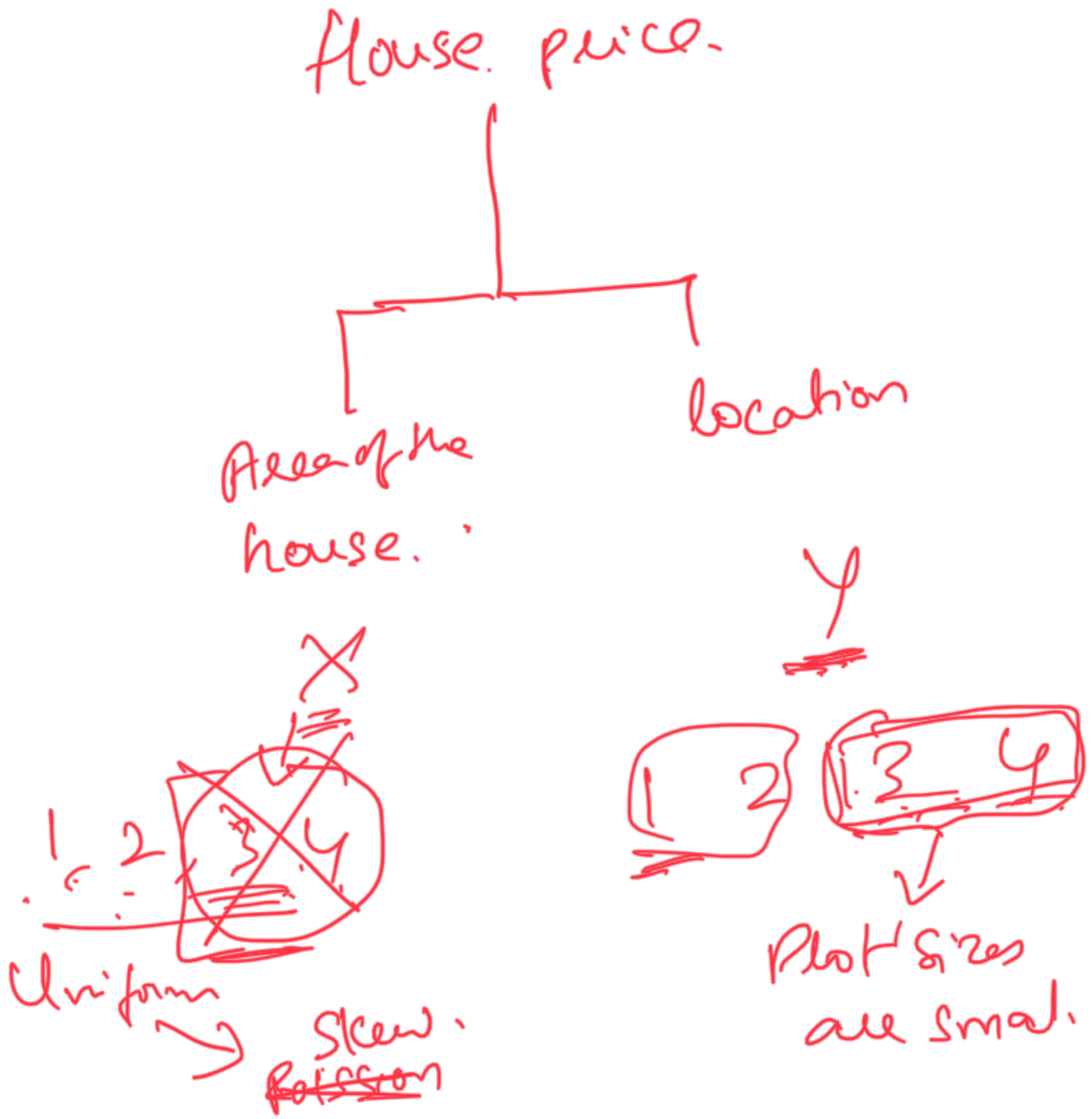
$$I(X;Y) = H(X) - H(X|Y)$$

→ H(X)

→ Conditional entropy

Independent $I(X:Y) = 0$


P=0.1
PMF
P=0.6
3
Y
2
1
X
X cannot take these values.


House price.
Area of the house.
location
X
Y
1 2 3 4
Uniform
Skew.
~~Poisson~~
1 2 3 4
Plot sizes are small.