Probability and Statistics

UG2, Core course, IIIT,H

Pawan Kumar

IIIT, Hyderabad

September 17, 2021

1 Probability and Information

1 Outline

→ Karmarkar (Interior Point Method)

In his 1948 paper "A Mathematical Theory of Communication", Claude Shannon introduced the revolutionary notion of Information Entropy.

Claude Shannon

Figure: Claude Shannon

American

AT&T Bell labs

→ Claud Shann (Communication theory)

Worst case

$\sqrt{N}$

→ Luv Grover (BTech BITS, Pilani ECE)

(Quantum Algo: Searching)

→ Schor's (Quantum Cryptography)

→ 6, 1, 2, 3, 4

O(N)

1 Entropy in Physics

1 Entropy in Physics

Entropy: Log of the number of microstates and microscopic configurations.

1 Entropy in Physics

Entropy: Log of the number of microstates and microscopic configurations.

1. if particles in a system have many possible positions to move, then the system has high entropy

1 Entropy in Physics

Entropy: Log of the number of microstates and microscopic configurations.

1. if particles in a system have many possible positions to move, then the system has high entropy
2. if they have to stay rigid, then the system has low entropy

Entropy: Log of the number of microstates and microscopic configurations.

1. if particles in a system have many possible positions to move, then the system has high entropy
2. if they have to stay rigid, then the system has low entropy

For example,

1. The molecules in ice have to stay in a lattice, as it is a rigid system, so ice has low entropy

1 Entropy in Physics

Entropy: Log of the number of microstates and microscopic configurations.

1. if particles in a system have many possible positions to move, then the system has high entropy
2. if they have to stay rigid, then the system has low entropy

For example,

1. The molecules in ice have to stay in a lattice, as it is a rigid system, so ice has low entropy
2. Molecules in water have more positions to move around, so water in liquid has medium entropy

1 Entropy in Physics

Entropy: Log of the number of microstates and microscopic configurations.

1. if particles in a system have many possible positions to move, then the system has high entropy
2. if they have to stay rigid, then the system has low entropy

For example,

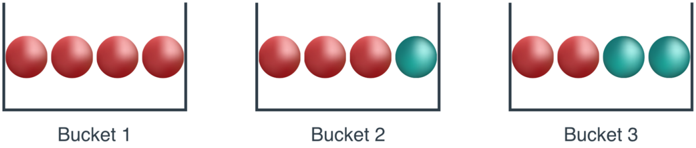
1. The molecules in ice have to stay in a lattice, as it is a rigid system, so ice has low entropy
2. Molecules in water have more positions to move around, so water in liquid has medium entropy
3. Molecules inside water vapor can pretty much go anywhere they want, so it has high entropy

Figure: Water in three phases

1 Entropy and Knowledge

1 Entropy and Knowledge

Shannon



2. In the second bucket, we know with 75% certainty that the ball is red.

Figure: The buckets

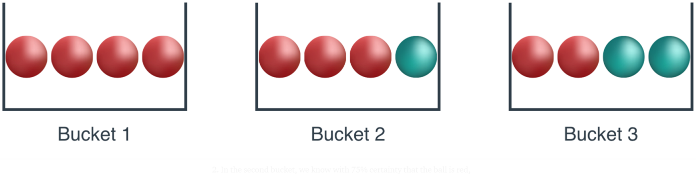


Figure: The buckets

1. In the first bucket, we'll know for sure that the ball coming out is red

1 Entropy and Knowledge

$\frac{3}{4} = 0.75$ 75%.

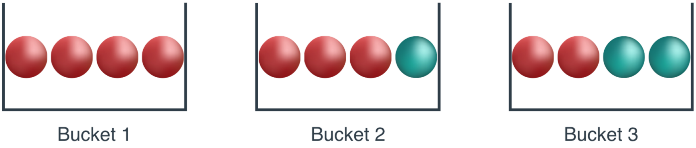


Figure: The buckets

1. In the first bucket, we'll know for sure that the ball coming out is red
2. In the second bucket, we know with 75% certainty that the ball is red, and with 25% certainty that it's blue

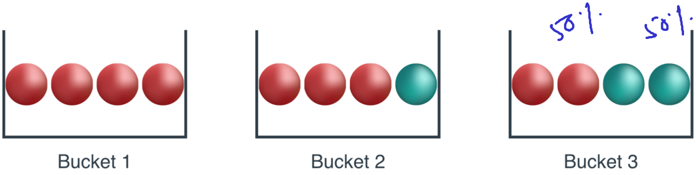


Figure: The buckets

1. In the first bucket, we'll know for sure that the ball coming out is red
2. In the second bucket, we know with 75% certainty that the ball is red, and with 25% certainty that it's blue
3. In the third bucket, we know with 50% certainty that the ball is red, and with the same certainty that it's blue

1 Entropy in buckets

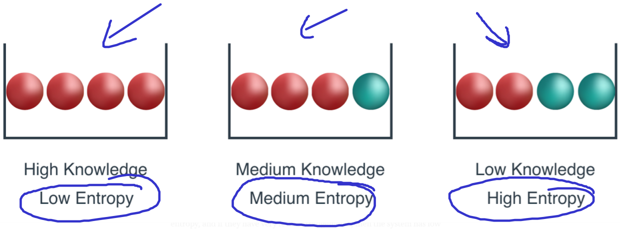


Figure: Entropy in buckets

1 Entropy in buckets

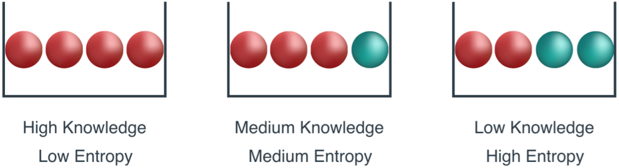


Figure: Entropy in buckets

1. Bucket 1 gives us the most amount of “knowledge” about what ball we’ll draw (because we know for sure it’s red)

1 Entropy in buckets

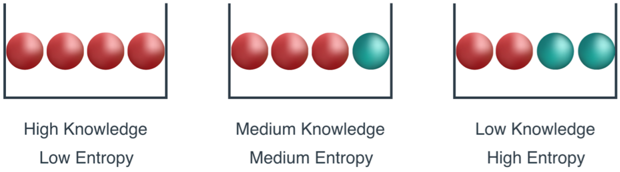


Figure: Entropy in buckets

1. Bucket 1 gives us the most amount of "knowledge" about what ball we'll draw (because we know for sure it's red)
2. Bucket 2 gives us some knowledge

1 Entropy in buckets

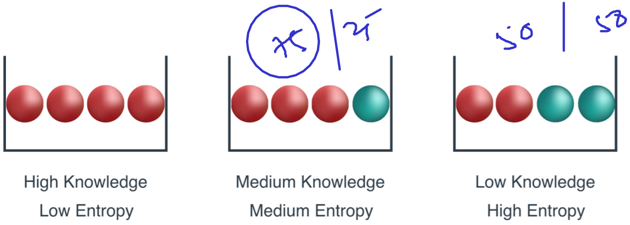


Figure: Entropy in buckets

1. Bucket 1 gives us the most amount of "knowledge" about what ball we'll draw (because we know for sure it's red)
2. Bucket 2 gives us some knowledge
3. Bucket 3 will give us the least amount of knowledge

"Entropy is in some way, the opposite of knowledge"

1 How to Quantify Entropy? Using Probability!

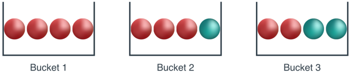


Figure: The buckets

Right Question to Ask

How do we cook up a formula which gives us a low number for a bucket with 4 red balls, a high number for a bucket with 2 red and 2 blue balls, and a medium number for a bucket with 3 red and 1 blue balls?

1 How to Quantify Entropy? Using Probability!

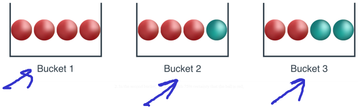


Figure: The buckets

Right Question to Ask

How do we cook up a formula which gives us a low number for a bucket with 4 red balls, a high number for a bucket with 2 red and 2 blue balls, and a medium number for a bucket with 3 red and 1 blue balls?

Recall Physical Entropy

If molecules have many possible rearrangements, then the system has high entropy, and if they have very few rearrangements, then the system has low entropy.

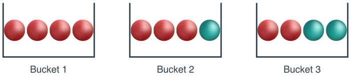


Figure: The buckets

Right Question to Ask

How do we cook up a formula which gives us a low number for a bucket with 4 red balls, a high number for a bucket with 2 red and 2 blue balls, and a medium number for a bucket with 3 red and 1 blue balls?

Recall Physical Entropy

If molecules have many possible rearrangements, then the system has high entropy, and if they have very few rearrangements, then the system has low entropy.

Possible Solution Using Counting

So a first attempt would be to count the number of rearrangements of these balls.

1 Entropy and Rearrangements

Figure: Entropy and Rearrangements

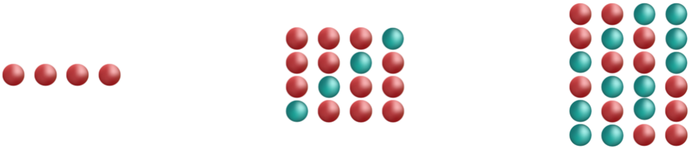


Figure: Entropy and Rearrangements

In this case, we have 1 possible rearrangement for Bucket 1, 4 for Bucket 2, and 6 for Bucket 3, this number given by the binomial coefficient.

Figure: Entropy and Rearrangements

In this case, we have 1 possible rearrangement for Bucket 1, 4 for Bucket 2, and 6 for Bucket 3, this number given by the binomial coefficient.

Hint of a formula for entropy!

If there are many arrangements, then entropy is large, and if there are very few arrangements, then entropy is low. In the next section, we'll cook up a formula for entropy.

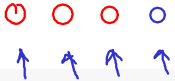
1 Entropy and an Interesting Game Show

Spoiler: The probability of winning this game, will help us get the formula for entropy.

Spoiler: The probability of winning this game, will help us get the formula for entropy.

1. We choose one of the three buckets. We are shown the balls in the bucket, in some order. Then, the balls go back in the bucket

1 Entropy and an Interesting Game Show



Spoiler: The probability of winning this game, will help us get the formula for entropy.

1. We choose one of the three buckets. We are shown the balls in the bucket, in some order. Then, the balls go back in the bucket
2. We then pick one ball out of the bucket, at a time, record the color, and return the ball back to the bucket

Spoiler: The probability of winning this game, will help us get the formula for entropy.

1. We choose one of the three buckets. We are shown the balls in the bucket, in some order. Then, the balls go back in the bucket
2. We then pick one ball out of the bucket, at a time, record the color, and return the ball back to the bucket
3. If the colors recorded make the same sequence than the sequence of balls that we were shown at the beginning, then we win 1,000,000 dollars. If not, then we lose

1 Game with an Example

$3|4$

Example

We're shown the balls in the bucket in some order, so let's say, they're shown to us in that precise order, red, red, red, blue. Now, let's try to draw the balls to get that sequence, red, red, red, blue. What's the probability of this happening?

Example

We're shown the balls in the bucket in some order, so let's say, they're shown to us in that precise order, red, red, red, blue. Now, let's try to draw the balls to get that sequence, red, red, red, blue. What's the probability of this happening?

1. In order for the first ball to be red, the probability is 3/4, or 0.75

Example

We're shown the balls in the bucket in some order, so let's say, they're shown to us in that precise order, red, red, red, blue. Now, let's try to draw the balls to get that sequence, red, red, red, blue. What's the probability of this happening?

1. In order for the first ball to be red, the probability is $3/4$, or 0.75
2. For the second ball to be red, the probability is again $3/4$. (Remember that we put the first ball back in the bucket after looking at its color.)

Example

We're shown the balls in the bucket in some order, so let's say, they're shown to us in that precise order, red, red, red, blue. Now, let's try to draw the balls to get that sequence, red, red, red, blue. What's the probability of this happening?

1. In order for the first ball to be red, the probability is 3/4, or 0.75
2. For the second ball to be red, the probability is again 3/4. (Remember that we put the first ball back in the bucket after looking at its color.)
3. For the third ball to be red, the probability is again 3/4

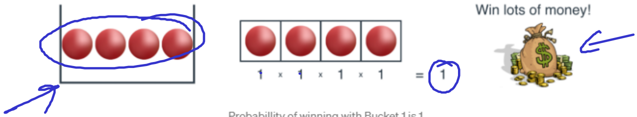
$P(R_1 \cap R_2 \cap R_3 \cap B_4) = P(R)P(R)P(R)P(B)$

Example

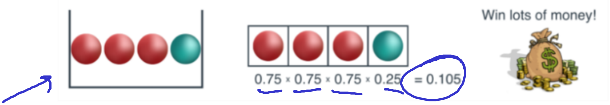
We're shown the balls in the bucket in some order, so let's say, they're shown to us in that precise order, red, red, red, blue. Now, let's try to draw the balls to get that sequence, red, red, red, blue. What's the probability of this happening?

1. In order for the first ball to be red, the probability is 3/4, or 0.75
2. For the second ball to be red, the probability is again 3/4. (Remember that we put the first ball back in the bucket after looking at its color.) 0. 1
3. For the third ball to be red, the probability is again 3/4
4. For the fourth ball to be blue, the probability is now 1/4, or 0.25

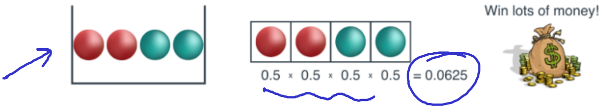
As these are independent events, then the probability of the 4 of them to happen, is $(3/4)*(3/4)*(3/4)*(1/4) = 27/256$, or 0.105. This is not very likely.



Probabillity of winning with Bucket 1 is 1



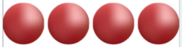
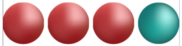
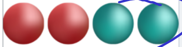
Probabillity of winning with Bucket 2 is 0.105

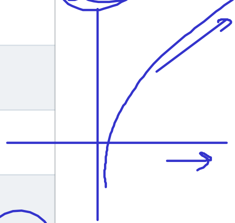


Probabillity of winning with Bucket 3 is 0.0625

1 Probability of Winning

$\times 10^{-16} \approx 0$

	P(red)	P(blue)	P(winning)
	1	0	1 × 1 × 1 × 1 = **1**
	0.75	0.25	0.75 × 0.75 × 0.75 × 0.25 = **0.105**
	0.5	0.5	0.5 × 0.5 × 0.5 × 0.5 = **0.0625**



$0.5 \times 4 = 2$

Towards Entropy Formula

In order to build the entropy formula, we want the opposite, some measure that gives us a low number for Bucket 1, a medium number for Bucket 2, and a high number for Bucket 3. No problem, this is where logarithms will come to save our life.

1 Calculate Entropy for Example

1 Calculate Entropy for Example

1 For bucket 2, (3 red balls, and 1 blue balls)

1 Calculate Entropy for Example

1 For bucket 2, (3 red balls, and 1 blue balls)

$$0.75 * 0.75 * 0.75 * 0.25 = 0.\underline{1054687}5$$

1 Calculate Entropy for Example

1. For bucket 2, (3 red balls, and 1 blue balls)

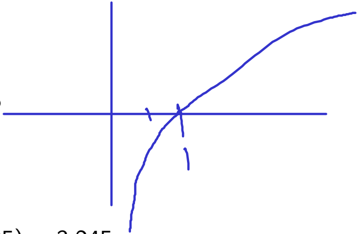
$$0.75 * 0.75 * 0.75 * 0.25 = 0.10546875$$

2. Taking the logarithm and multiplying by -1, we get

1 Calculate Entropy for Example

1 For bucket 2, (3 red balls, and 1 blue balls)

$$0.75 * 0.75 * 0.75 * 0.25 = 0.10546875$$



2 Taking the logarithm and multiplying by -1, we get

$$-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25) = 3.245$$

1 Calculate Entropy for Example

1. For bucket 2, (3 red balls, and 1 blue balls)

$$0.75 * 0.75 * 0.75 * 0.25 = 0.10546875$$

2. Taking the logarithm and multiplying by -1, we get

$$-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25) = 3.245$$

3. Finally, we take the average in order to normalize

1 Calculate Entropy for Example

1. For bucket 2, (3 red balls, and 1 blue balls)

$$0.75 * 0.75 * 0.75 * 0.25 = 0.10546875$$

2. Taking the logarithm and multiplying by -1, we get

$$-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25) = 3.245$$

3. Finally, we take the average in order to normalize

$$\frac{1}{4}(-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25)) = 0.81125$$

1 Calculate Entropy for Example

$1 \times 1 \times 1 \times 1 = 1$

$\log_2(\ ) = 0$

1. For bucket 2, (3 red balls, and 1 blue balls)

$$0.75 * 0.75 * 0.75 * 0.25 = 0.10546875$$

2. Taking the logarithm and multiplying by -1, we get

$$-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25) = 3.245$$

3. Finally, we take the average in order to normalize

$$\frac{1}{4}(-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25)) = 0.81125$$

4. Similarly, entropy for bucket 1: 0

1 Calculate Entropy for Example

1. For bucket 2, (3 red balls, and 1 blue balls)

$$0.75 * 0.75 * 0.75 * 0.25 = 0.10546875$$

2. Taking the logarithm and multiplying by -1, we get

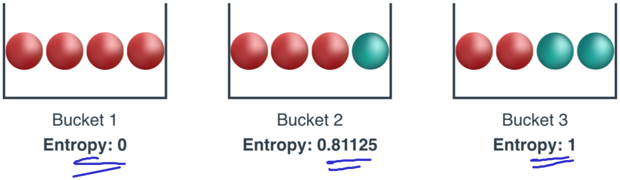
$$-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25) = 3.245$$

3. Finally, we take the average in order to normalize

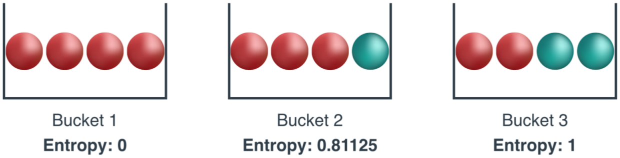
$$\frac{1}{4}(-\log_2(0.75) - \log_2(0.75) - \log_2(0.75) - \log_2(0.25)) = 0.81125$$

4. Similarly, entropy for bucket 1: 0
5. Similarly, entropy for bucket 2: 1

1 Entropy for m red balls and n blue balls



1 Entropy for *m* red balls and *n* blue balls



$$\text{Entropy} = \frac{-m}{m+n}\log_2\left(\frac{m}{m+n}\right) + \frac{-n}{m+n}\log_2\left(\frac{n}{m+n}\right)$$

General formula for Entropy

1 Multiclass Entropy

$4/8, \; 2/8, \; 1/8, \; 1/8$

AAAAAAAA	AAAABBCD	AABBCCDD
Bucket 1	Bucket 2	Bucket 3
Low Entropy	**Medium Entropy**	**High Entropy**

$2/8, \; 2/8, \; 2/8, \; 2/8$

Formula for Entropy

The general formula for entropy is

$$\text{Entropy} = -\sum_{i=1}^{n} p_i \log_2 p_i ,$$

where n is the number of classes, p_i is the probability of an object from the ith class appearing.

1 Entropy for the Buckets Using Formula

$$-\sum_{i=1}^{n} p_i \log_2 p_i$$

$$1 \cdot \log_2 1 = 0$$

$$\text{Entropy} = -1\log_2(1) = 0$$

Entropy for Bucket 1

$$\text{Entropy} = -\frac{4}{8}\log_2\left(\frac{4}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = 1.75$$

$$\text{Entropy} = -\frac{2}{8}\log_2\left(\frac{2}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) - \frac{2}{8}\log_2\left(\frac{2}{8}\right) = 2$$

AAAAAAAA	AAAABBCD	AABBCCDD
Bucket 1	Bucket 2	Bucket 3
Entropy = 0	**Entropy = 1.75**	**Entropy = 2**

Bucket 1	Bucket 2	Bucket 3
AAAAAAAA	AAAABBCD	AABBCCDD

On average, how many questions do we need to ask to find out what letter it is?

AAAAAAAA — Bucket 1

AAAABBCD — Bucket 2

AABBCCDD — Bucket 3

On average, how many questions do we need to ask to find out what letter it is?

1. For bucket 1: Avrage number of questions $= 0$

AAAAAAAA	AAAABBCD	AABBCCDD
Bucket 1	Bucket 2	Bucket 3

On average, how many questions do we need to ask to find out what letter it is?

1. For bucket 1: Avrage number of questions $= 0$
2. For buckets 2 and 3, we may think we need the following questions

1 Information Theory

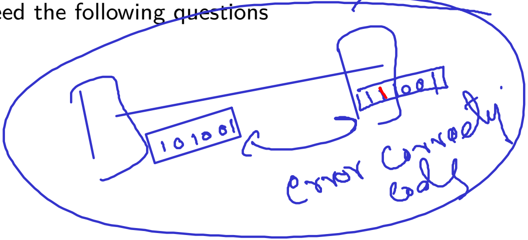
"Coding Theory"

"Information Theory"

AAAAAAAA	AAAABBCD	AABBCCDD
Bucket 1	Bucket 2	Bucket 3

On average, how many questions do we need to ask to find out what letter it is?

1. For bucket 1: Avrage number of questions = 0
2. For buckets 2 and 3, we may think we need the following questions
    1. Is the letter an A?
    2. Is the letter a B?
    3. Is the letter a C?
    4. ~~Is the letter a D?~~



AAAAAAAA AAAABBCD AABBCCDD

Bucket 1 Bucket 2 Bucket 3

On average, how many questions do we need to ask to find out what letter it is?

1. For bucket 1: Avrage number of questions $= 0$
2. For buckets 2 and 3, we may think we need the following questions
   1. Is the letter an A?
   2. Is the letter a B?
   3. Is the letter a C?
   4. Is the letter a D?

Do we need all the four questions? Is the last question needed? Can we do better?

AAAAAAAA AAAABBCD AABBCCDD

Bucket 1 Bucket 2 Bucket 3

On average, how many questions do we need to ask to find out what letter it is?

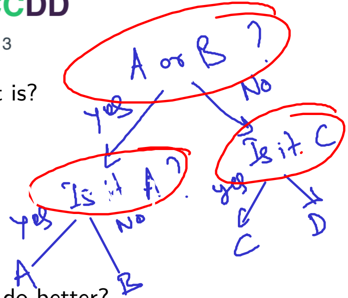
1. For bucket 1: Avrage number of questions $= 0$
2. For buckets 2 and 3, we may think we need the following questions
   1. Is the letter an A?
   2. Is the letter a B?
   3. Is the letter a C?
   4. Is the letter a D?

   Do we need all the four questions? Is the last question needed? Can we do better?

Consider the following questions:

AAAAAAAA AAAABBCD AABBCCDD

Bucket 1 Bucket 2 Bucket 3



On average, how many questions do we need to ask to find out what letter it is?

1. For bucket 1: Avrage number of questions = 0
2. For buckets 2 and 3, we may think we need the following questions
    1. Is the letter an A?
    2. Is the letter a B?
    3. Is the letter a C?
    4. Is the letter a D?

    Do we need all the four questions? Is the last question needed? Can we do better?

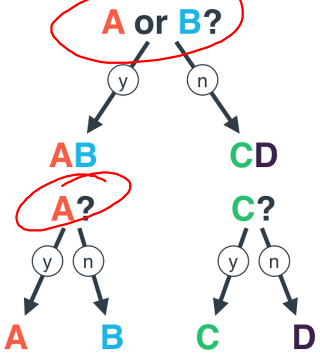
Consider the following questions:

1. Is the letter A or B?
    1. a) If the answer to question 1 is "yes": Is the letter A? If the answer to question 1 is "no": Is the letter C?

1 Average Number of Questions for Bucket 3

A A BB C C B D

$\frac{2}{8}\cdot 2 + \frac{2}{8}\cdot 2 + \frac{2}{8}\cdot 2 + \frac{2}{8}\cdot 2$



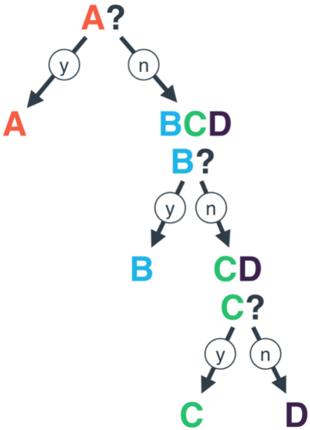
questions

$$\text{Average Number of Questions} = \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 = 2$$

1 Average Number of Questions for Bucket 2

AAAABBCD

$$\frac{4}{8}\cdot 1 + \frac{2}{8}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 = 1.75$$



$$\text{Average Number of Questions} = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 = 1.75$$

1 Conclusion

Here's the connection between Entropy and Information Theory. If we want to find out a letter drawn out of a bucket, the average number of questions we must ask to find out (if we ask our questions in the smartest possible way), is at least the entropy of the set.

1 Conclusion

Here's the connection between Entropy and Information Theory. If we want to find out a letter drawn out of a bucket, the average number of questions we must ask to find out (if we ask our questions in the smartest possible way), is at least the entropy of the set.

The entropy of the set is a lower bound on the number of questions we must ask in average to find out. In the cases we saw above, the number of questions is exactly the entropy.

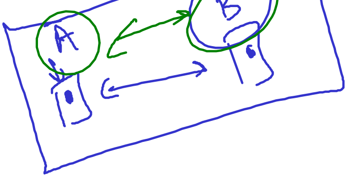
1 Conclusion

Here's the connection between Entropy and Information Theory. If we want to find out a letter drawn out of a bucket, the average number of questions we must ask to find out (if we ask our questions in the smartest possible way), is at least the entropy of the set.

"Nielson & Chuang"

The entropy of the set is a lower bound on the number of questions we must ask in average to find out. In the cases we saw above, the number of questions is exactly the entropy.

We may need to ask more questions than the entropy. But we will never be able to do it with less questions than the entropy of the set.



Claude Shannon

"The mathematics of Communication"