

# COMP2610 / COMP6261 Information Theory

## Lecture 2: First Steps and Basic Probability

**Thushara Abhayapala**

Audio & Acoustic Signal Processing Group  
School of Engineering,  
College of Engineering & Computer Science  
The Australian National University,  
Canberra, Australia.

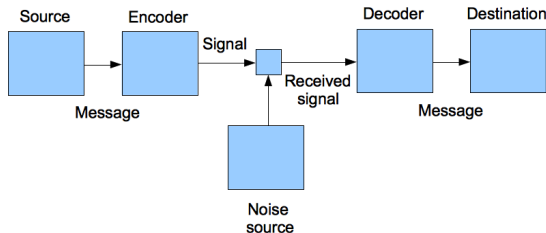


Australian  
National  
University

# Outline

- 1 A General Communication System
- 2 The Role of Uncertainty
- 3 Basic Concepts In Probability
- 4 Relating Joint, Marginal and Conditional Probabilities
- 5 Wrapping Up

# A General Communication System



**Source** : The information source that generates the message to be communicated

**Encoder** : Operates on the message to produce a signal suitable for transmission

**Channel** : The medium used to transmit the signal

**Decoder** : Reconstructs the message from the signal

**Destination** : The entity that receives the message

# Communication over Noisy Channels

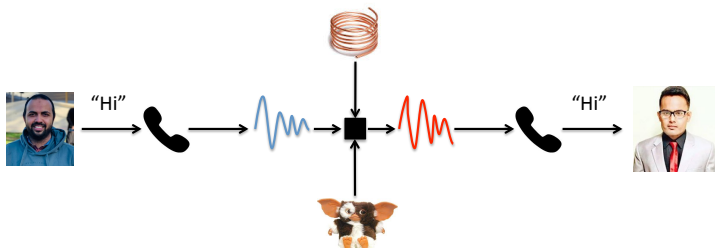
A **channel** is some medium for transmitting messages

A **noisy channel** is a channel that potentially introduces **errors** in the sender's message

## The Problem of Communication

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.”  
(Claude Shannon, 1948)

# Example: Telephone Network



Source : Akram

Encoder : Telephone handset

Channel : Analogue telephone line

Decoder : Telephone handset

Destination : Manish

# Examples of Noisy Communication Channels

Other examples of noisy channels:

- A radio communication link
- VDSL NBN connection
- The complete link from camera, through editing, Netflix, the rest of the internet, VDSL link to home, wifi to TV screen
- Reproducing cells
- A magnetic hard disk drive
  - ▶ Channel does not need to involve physical movement

What would the other components be for each of these channels?

- 1 A General Communication System
- 2 The Role of Uncertainty
- 3 Basic Concepts In Probability
- 4 Relating Joint, Marginal and Conditional Probabilities
- 5 Wrapping Up

# Uncertainty in Communication

We can not avoid **uncertainty** when

- 1 Dealing with noise (imperfections) in the channel
- 2 “Compressing” the messages (compare a high-resolution photograph of a manuscript with the typed text that captures the essence; or a transcript of a spoken utterance)



# Channel Noise

A noisy channel introduces errors in sender's message

Thus, receiver is **uncertain** that the message is what the sender intended

How to model, quantify, and mitigate this uncertainty?

# Message Compression – I

Cover and Thomas, Example 1.1.2

- Suppose we'd like to relay the outcome of an 8 horse race to a friend
  - ▶ We wish to convey one of  $\{ A, B, \dots, H \}$
- Suppose we encode the message as a binary string. A natural coding scheme is

A  $\rightarrow$  000

B  $\rightarrow$  001

C  $\rightarrow$  010

$\vdots$

H  $\rightarrow$  111

# Message Compression – II

Cover and Thomas, Example 1.1.2

- Now say the probabilities of the horses winning are  $(1/2, 1/4, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64)$
- Encoding messages based on their **probability** of the being chosen will give shorter **expected** lengths:

A  $\rightarrow$  0

B  $\rightarrow$  10

C  $\rightarrow$  110

D  $\rightarrow$  1110

E  $\rightarrow$  11110

F  $\rightarrow$  111100

G  $\rightarrow$  111101

H  $\rightarrow$  111111

# What is “Information”?

For noise correction and message compression, we will need to quantify the **information** contained in a message

Roughly, “information” measures how much:

- **Unexpected** data a message contains
- The receiver's **uncertainty** is reduced on seeing the message

# The Case for Probability

We run into the notion of **uncertainty** when trying to pin down:

- 1 How to deal with channel noise
- 2 How to compress messages
- 3 What “information” means

To make progress, we need to formalise uncertainty

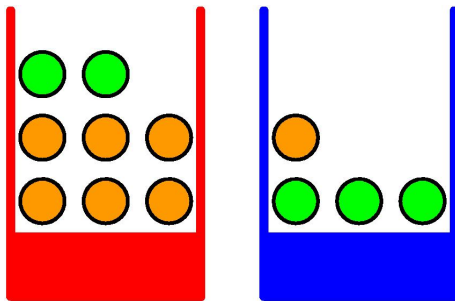
We will do this using **probability theory**

We now commence our *review* of probability; this will be hard going if you have not met it before!

- 1 A General Communication System
- 2 The Role of Uncertainty
- 3 Basic Concepts In Probability**
- 4 Relating Joint, Marginal and Conditional Probabilities
- 5 Wrapping Up

# Probability: Example

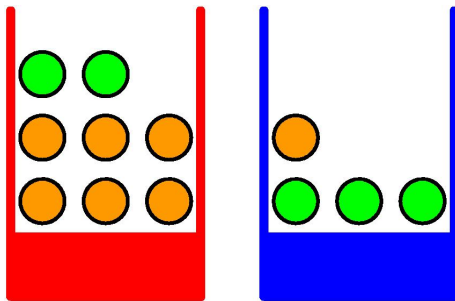
Quantification and Manipulation of Uncertainty (Bishop, PRML, 2006)



# Probability: Example

Quantification and Manipulation of Uncertainty (Bishop, PRML, 2006)

- 1 Pick a box at random

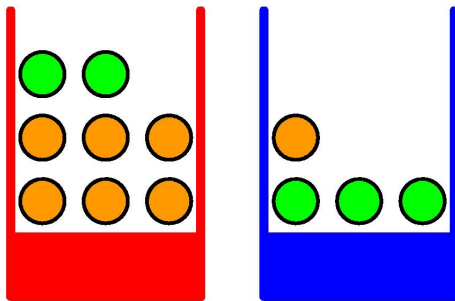




# Probability: Example

Quantification and Manipulation of Uncertainty (Bishop, PRML, 2006)

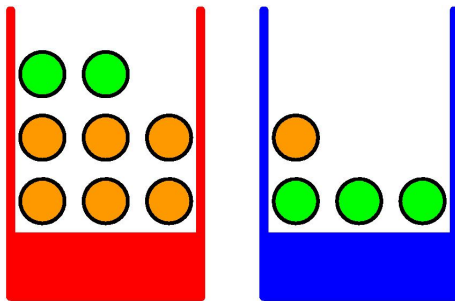
- 1 Pick a box at random
- 2 From the selected box, pick a fruit (apple or orange) at random



# Probability: Example

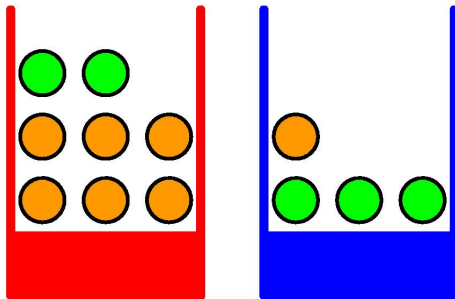
Quantification and Manipulation of Uncertainty (Bishop, PRML, 2006)

- 1 Pick a box at random
- 2 From the selected box, pick a fruit (apple or orange) at random
- 3 Observe the fruit type and return it back to the original box



# Probability: Example

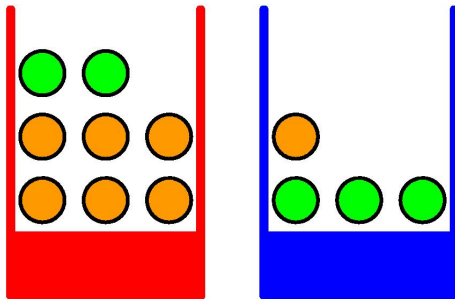
Quantification and Manipulation of Uncertainty (Bishop, PRML, 2006) — Cont'd



- Identity of the box is a random variable  $B \in \{r, b\}$
- Identity of the fruit is a random variable  $F \in \{a, o\}$

# Probability: Example

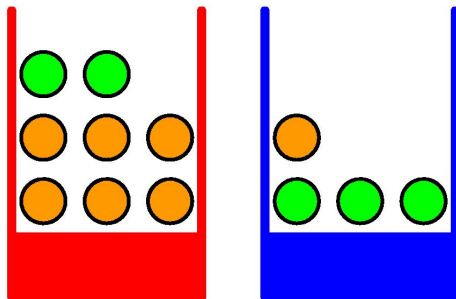
Quantification and Manipulation of Uncertainty (Bishop, PRML, 2006) — Cont'd



- Identity of the box is a random variable  $B \in \{r, b\}$
- Identity of the fruit is a random variable  $F \in \{a, o\}$

**Probability of an event:** Proportion of times it happens out of a large number of trials

## Probability: Example



Suppose we repeat the selection process many times, and ended up picking up the blue box 60% of the time and the red box 40% of the time

- $p(B = r) = 0.4, p(B = b) = 0.6$

# Probability: Basic Properties

By definition,  $0 \leq p(B = b) \leq 1$  and  $0 \leq p(B = r) \leq 1$

Outcomes are mutually exclusive:

$$\begin{aligned} p(B = r \text{ AND } B = b) &= p(B = r, B = b) \\ &= 0 \end{aligned}$$

Outcomes are jointly exhaustive:

$$\begin{aligned} p(B = r \text{ OR } B = b) &= p(B = r) + p(B = b) - p(B = r \text{ AND } B = b) \\ &= p(B = r) + p(B = b) \\ &= 1 \end{aligned}$$

# Probability

## What Types of Questions Can We Answer?

- What is the probability of picking the red box, and an apple within that box?
- What is the (overall) probability of picking up an apple?
- Given that we selected a red box, what is the probability of selecting an apple?

We can answer these and more complex questions by using the *rules of probability*.

## Joint Probability

What is the probability of selecting the red box **and** selecting an apple?



# Joint Probability

What is the probability of selecting the red box **and** selecting an apple?

## Joint Probability of a Set of Events

The proportion of times these events happened **together** out of the total number of trials.

# Joint Probability

What is the probability of selecting the red box **and** selecting an apple?

## Joint Probability of a Set of Events

The proportion of times these events happened **together** out of the total number of trials.

If we repeated our experiment many (say  $N = 100$ ) times, and in 10 of the trials we saw  $B = r$  and  $F = a$ , then we may estimate

$$\begin{aligned} p(B = r \text{ AND } F = a) &= p(B = r, F = a) \\ &= \frac{10}{100} \\ &= 0.1 \end{aligned}$$

## Marginal Probability

What is the probability of an apple being picked, **regardless** of the box we selected?

# Marginal Probability

What is the probability of an apple being picked, **regardless** of the box we selected?

## Marginal Probability of an Event

The proportion of times that this event happened out of the total number of trials.

# Marginal Probability

What is the probability of an apple being picked, **regardless** of the box we selected?

## Marginal Probability of an Event

The proportion of times that this event happened out of the total number of trials.

Remember that we selected a red box and an apple in 10 out of 100 trials

Say that in 45 of the trials, we selected a blue box and an apple

So, **irrespective of  $B$** , an apple was selected  $45 + 10 = 55$  times, and:

$$p(F = a) = \frac{55}{100} = \frac{11}{20}$$

## Conditional Probability

What is the probability of an apple being picked up, **given that** a red box was selected?

# Conditional Probability

What is the probability of an apple being picked up, **given that** a red box was selected?

## Conditional Probability of an Event

The conditional probability of an event  $X$  with respect to an event  $Y$  is the proportion of times that  $X$  happens out of the times that  $Y$  happens.

# Conditional Probability

What is the probability of an apple being picked up, **given that** a red box was selected?

## Conditional Probability of an Event

The conditional probability of an event  $X$  with respect to an event  $Y$  is the proportion of times that  $X$  happens out of the times that  $Y$  happens.

The trials where we selected a blue box are **irrelevant**, whether or not an apple was selected

We selected a red box **and** an apple 10 out of 100 times

We selected a red box (regardless of the fruit) 40 out of 100 times

$$p(F = a \text{ GIVEN } B = r) = p(F = a | B = r) = \frac{10}{40} = \frac{1}{4}$$



# Conditional Probability

What is the probability of an apple being picked up, **given that** a red box was selected?

## Conditional Probability of an Event

The conditional probability of an event  $X$  with respect to an event  $Y$  is the proportion of times that  $X$  happens out of the times that  $Y$  happens.

The trials where we selected a blue box are **irrelevant**, whether or not an apple was selected

We selected a red box **and** an apple 10 out of 100 times

We selected a red box (regardless of the fruit) 40 out of 100 times

$$p(F = a \text{ GIVEN } B = r) = p(F = a | B = r) = \frac{10}{40} = \frac{1}{4}$$

Can we write this in terms of the joint and marginal probabilities?

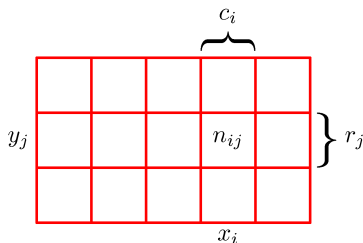
- 1 A General Communication System
- 2 The Role of Uncertainty
- 3 Basic Concepts In Probability
- 4 Relating Joint, Marginal and Conditional Probabilities**
- 5 Wrapping Up

# Joint, Marginal and Conditional Probabilities:

## A More General Formulation (1)

Consider the more general case of two random variables:

$$X \in \{x_1, x_2, \dots, x_M\} \text{ and } Y \in \{y_1, y_2, \dots, y_L\}$$



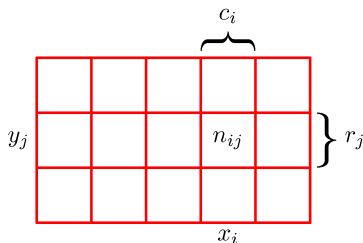
$N$  : Total number of trials

# Joint, Marginal and Conditional Probabilities:

## A More General Formulation (1)

Consider the more general case of two random variables:

$$X \in \{x_1, x_2, \dots, x_M\} \text{ and } Y \in \{y_1, y_2, \dots, y_L\}$$



$N$  : Total number of trials

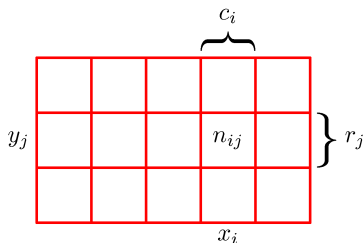
$n_{ij}$  :  $\#(X = x_i, Y = y_j) = \#$  of times that  $x_i$  and  $y_j$  happen

# Joint, Marginal and Conditional Probabilities:

## A More General Formulation (1)

Consider the more general case of two random variables:

$$X \in \{x_1, x_2, \dots, x_M\} \text{ and } Y \in \{y_1, y_2, \dots, y_L\}$$



$N$  : Total number of trials

$n_{ij}$  :  $\#(X = x_i, Y = y_j) = \#$  of times that  $x_i$  and  $y_j$  happen

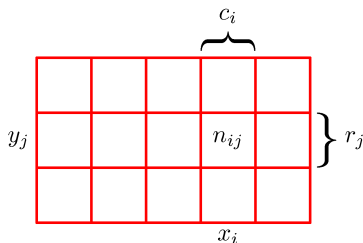
$c_i$  :  $\#(X = x_i) = \sum_j n_{ij} = \#$  of times that  $x_i$  happens

# Joint, Marginal and Conditional Probabilities:

## A More General Formulation (1)

Consider the more general case of two random variables:

$$X \in \{x_1, x_2, \dots, x_M\} \text{ and } Y \in \{y_1, y_2, \dots, y_L\}$$



$N$  : Total number of trials

$n_{ij}$  :  $\#(X = x_i, Y = y_j) = \#$  of times that  $x_i$  and  $y_j$  happen

$c_i$  :  $\#(X = x_i) = \sum_j n_{ij} = \#$  of times that  $x_i$  happens

$r_j$  :  $\#(Y = y_j) = \sum_i n_{ij} = \#$  of times that  $y_j$  happens

# Joint, Marginal and Conditional Probabilities:

## A More General Formulation (2)

			$n_{ij}$	

By definition:

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} \text{ (Joint)}$$

$$p(X = x_i) = \frac{c_i}{N} \text{ (Marginal)}$$

$$p(Y = y_j) = \frac{r_j}{N} \text{ (Marginal)}$$

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i} \text{ (Conditional)}$$

# Joint, Marginal and Conditional Probabilities:

## A More General Formulation (1)

Bins and fruit example:

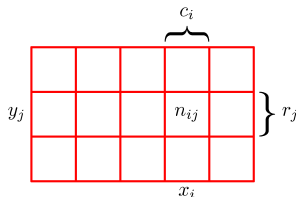
	<b>Orange</b>	<b>Apple</b>
<b>Blue</b>	15	45
<b>Red</b>	30	10

Verify the computations from previous section with this table



# Joint, Marginal and Conditional Probabilities:

## A More General Formulation (3)



Observe:

$$\begin{aligned} p(X = x_i) &= \frac{\sum_j n_{ij}}{N} \\ &= \sum_j p(X = x_i, Y = y_j) \end{aligned}$$

$$\begin{aligned} p(Y = y_j | X = x_i) &= \frac{n_{ij}}{c_i} = \frac{n_{ij}}{N} \bigg/ \frac{c_i}{N} \\ &= p(X = x_i, Y = y_j) / p(X = x_i) \end{aligned}$$

# The Rules of Probability

Sum Rule / Marginalization :

$$p(X = x_i) = \sum_j p(X = x_i, Y = y_j)$$

# The Rules of Probability

Sum Rule / Marginalization :

$$p(X = x_i) = \sum_j p(X = x_i, Y = y_j)$$

Product Rule :

$$p(X = x_i, Y = y_j) = p(Y = y_j | X = x_i) p(X = x_i)$$

# The Rules of Probability

Sum Rule / Marginalization :

$$p(X = x_i) = \sum_j p(X = x_i, Y = y_j)$$

Product Rule :

$$p(X = x_i, Y = y_j) = p(Y = y_j|X = x_i)p(X = x_i)$$

and by symmetry:

$$P(Y = y_j, X = x_i) = p(X = x_i|Y = y_j)p(Y = y_j)$$

# The Rules of Probability

Sum Rule / Marginalization :

$$p(X = x_i) = \sum_j p(X = x_i, Y = y_j)$$

Product Rule :

$$p(X = x_i, Y = y_j) = p(Y = y_j | X = x_i) p(X = x_i)$$

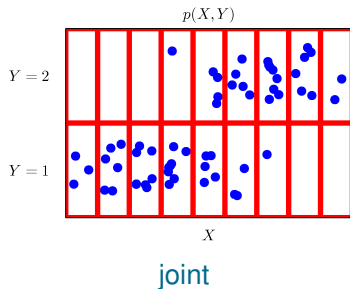
and by symmetry:

$$P(Y = y_j, X = x_i) = p(X = x_i | Y = y_j) p(Y = y_j)$$

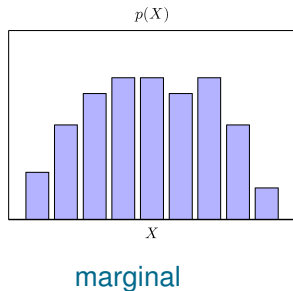
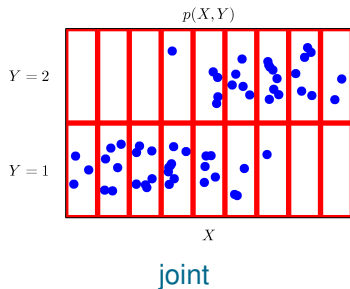
Therefore:

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j P(X = x_i | Y = y_j) P(Y = y_j)$$

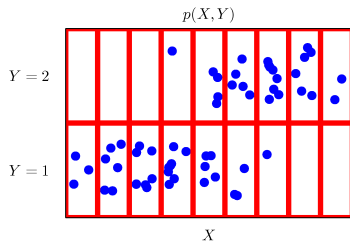
# An Illustration of a Distribution over Two Variables



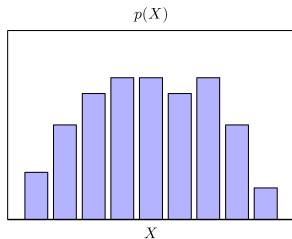
# An Illustration of a Distribution over Two Variables



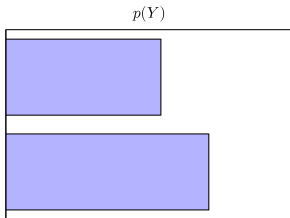
# An Illustration of a Distribution over Two Variables



joint



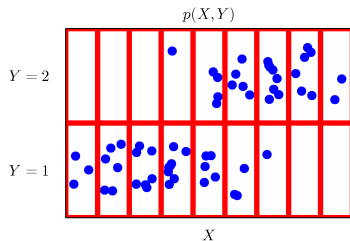
marginal



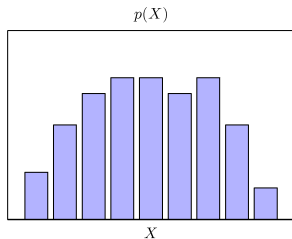
marginal



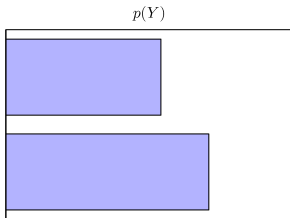
# An Illustration of a Distribution over Two Variables



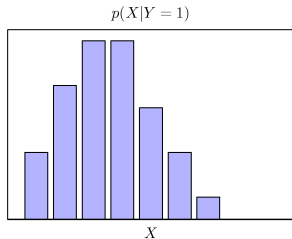
joint



marginal



marginal



conditional

# Joint, Marginal and Conditional Probabilities:

An even More General Formulation

Given  $D$  random variables  $X_1, \dots, X_D$ :

$$p(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_D) = \sum_{X_i} p(X_1, \dots, X_D)$$

# Joint, Marginal and Conditional Probabilities:

An even More General Formulation

Given  $D$  random variables  $X_1, \dots, X_D$ :

$$p(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_D) = \sum_{X_i} p(X_1, \dots, X_D)$$

**Chain Rule:** We can also express:

$$p(X_1, X_2) = p(X_1)p(X_2|X_1) \quad \text{What are we using here?}$$

# Joint, Marginal and Conditional Probabilities:

## An even More General Formulation

Given  $D$  random variables  $X_1, \dots, X_D$ :

$$p(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_D) = \sum_{X_i} p(X_1, \dots, X_D)$$

**Chain Rule:** We can also express:

$$p(X_1, X_2) = p(X_1)p(X_2|X_1) \quad \text{What are we using here?}$$

$$p(X_1, X_2, X_3) = p(X_1, X_2)p(X_3|X_1, X_2) = p(X_1)p(X_2|X_1)p(X_3|X_1, X_2)$$

# Joint, Marginal and Conditional Probabilities:

## An even More General Formulation

Given  $D$  random variables  $X_1, \dots, X_D$ :

$$p(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_D) = \sum_{X_i} p(X_1, \dots, X_D)$$

**Chain Rule:** We can also express:

$$p(X_1, X_2) = p(X_1)p(X_2|X_1) \quad \text{What are we using here?}$$

$$p(X_1, X_2, X_3) = p(X_1, X_2)p(X_3|X_1, X_2) = p(X_1)p(X_2|X_1)p(X_3|X_1, X_2)$$

$$p(X_1, \dots, X_D) = p(X_1)p(X_2|X_1)p(X_3|X_2, X_1) \dots p(X_D|X_1, \dots, X_{D-1})$$

- 1 A General Communication System
- 2 The Role of Uncertainty
- 3 Basic Concepts In Probability
- 4 Relating Joint, Marginal and Conditional Probabilities
- 5 Wrapping Up

# Summary

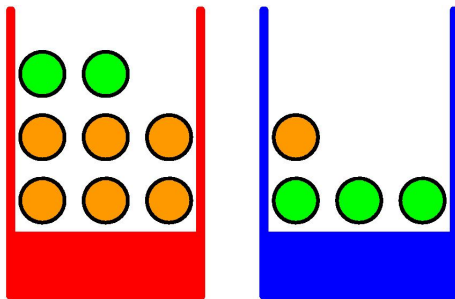
- General architecture for communication systems
- Why we need probability
- Probability theory: joint, marginal and conditional distribution
- **Reading:** Mackay § 2.1 and § 2.2; Bishop § 1.2

# Exercise

Coming Back to our Original Example

Given:  $p(B = r) = 0.4$ ,  $p(B = b) = 0.6$

Assume the fruit are selected uniformly from each box



- Write down all conditional probabilities  $P(F|B)$
- Evaluate the overall probabilities  $P(F)$



## Next time

- More on joint, marginal and conditional distributions
- When can we say that  $X$ ,  $Y$  do not influence each other?
- What, if anything, does  $p(X = x|Y = y)$  tell us about  $p(Y = y|X = x)$ ?

# Acknowledgement

These slides were originally developed by Professor Robert C. Williamson.