Channels

Mário S. Alvim
(msalvim@dcc.ufmg.br)

Quantitative Information Flow

DCC-UFMG
(2021/1)

Introduction

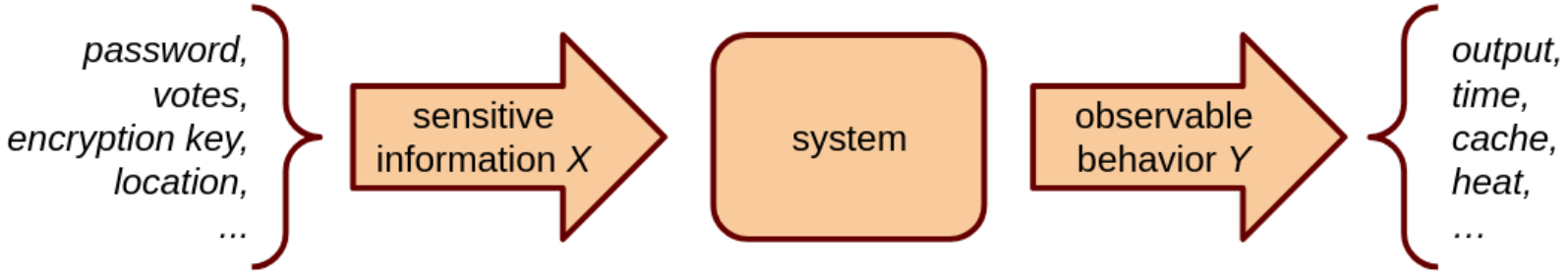
- We have so far considered secrets and how vulnerable they may be given an adversary's a priori knowledge.
- We'll now consider **systems** that process secret information.

  These systems are important because they are the means by which an adversary can update her knowledge about the secret.

  We model such systems as information-theoretic channels, which are probabilistic functions from inputs to outputs.

Introduction

- A **channel** is a system that:



1. takes as input a secret X, whose possible values come from a finite set $\mathcal{X}$; and
2. produces as only observable behavior an output Y, coming from a finite set $\mathcal{Y}$.

- To model a probabilistic system, we use a function

$$\mathsf{C} : \mathcal{X} \to \mathbb{D}\mathcal{Y}$$

mapping each $x \in \mathcal{X}$ to a distribution on $\mathcal{Y}$.

A special case is a deterministic system where each possible input gives rise to a unique output (i.e., $\mathsf{C} : \mathcal{X} \to \mathcal{Y}$).

Channel matrices

Channel matrices

- Channels can be described as information-theoretic channel matrices.

Definition (4.1)

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets, intuitively representing secret input values and observable output values.

A **channel matrix** C **from** $\mathcal{X}$ **to** $\mathcal{Y}$ is a matrix, indexed by $\mathcal{X} \times \mathcal{Y}$, whose rows give the distribution on outputs corresponding to each possible input.

That is, entry $\mathsf{C}_{x,y}$ denotes $p(y|x)$, the conditional probability of getting output y given input x.

Note that all entries in a channel matrix are between 0 and 1, and each row sums to 1, which property is called **stochastic** (and is how we will describe such matrices even if they might not correspond to actual channels).

The x'th row of C is written $\mathsf{C}_{x,-}$, and similarly the y'th column is $\mathsf{C}_{-,y}$.

Channel matrices

- Mathematically, a channel matrix has type

$$\mathcal{X} \times \mathcal{Y} \to [0,1] \qquad \text{or} \qquad \mathcal{X} \to \mathbb{D}\mathcal{Y} \ .$$

  But we will often prefer to write the type of a channel more concisely as

$$\mathcal{X} \to\!\!\!\!\triangleright\, \mathcal{Y} \ ,$$

  and thinking of it as a "probabilistic function" from $\mathcal{X}$ to $\mathcal{Y}$.

- Example 1 Here is a channel matrix C:

C	y_1	y_2	y_3	y_4
x_1	$1/2$	$1/2$	0	0
x_2	0	$1/4$	$1/2$	$1/4$
x_3	$1/2$	$1/3$	$1/6$	0

  For example, in C we have $p(y_3|x_3) = 1/6$. ◁

Channel matrices

- In general, channel matrices are (properly) probabilistic.

  However, an important special case is the deterministic channel matrix, identifying a unique possible output for each input.

- Example 2 Here is a deterministic channel matrix D:

D	y_1	y_2
x_1	1	0
x_2	1	0
x_3	0	1

  For example, in D we have $p(y_2|x_3) = 1$, meaning that whenever the input is x_3 the output is determined to be y_2. ◁

Channel matrices

- Channel matrices can be cumbersome to write down explicitly when $\mathcal{X}$ or $\mathcal{Y}$ are large.

  We sometimes find it convenient to represent channels as pseudocode.

- Example 3 A certain deterministic channel taking as input a 64-bit unsigned X (whose channel matrix hence has 2^{64} rows) can be described by the following pseudocode:

$$Y := \texttt{ if } (X \texttt{ mod } 8) = 0 \texttt{ then } X \texttt{ else } 1$$

◁

The effect of a channel on the adversary's knowledge

The effect of a channel on the adversary's knowledge

- We now consider how a channel affects the adversary's knowledge about a secret X with prior distribution π.

- Example 4 Suppose that channel matrix C below produces output y_4.

C	y_1	y_2	y_3	y_4
x_1	$1/2$	$1/2$	0	0
x_2	0	$1/4$	$1/2$	$1/4$
x_3	$1/2$	$1/3$	$1/6$	0

In that case, the adversary can deduce that X must be x_2, since that is the only input that can have led to that output. ◁

- We're relying on the assumption that she knows how C works:

Throughout this course, we make the worst-case assumption that the adversary knows how the channel works, in the sense of knowing the channel matrix C.

At a minimum, that means that she can compute $C_{x,y}$ for any $x : \mathcal{X}$ and $y : \mathcal{Y}$.

The effect of a channel on the adversary's knowledge

- In general, observing an output y allows the adversary to update her knowledge about X:
  - from the prior distribution π
  - to a posterior distribution, which we denote $p_{X|y}$.
- The relevant probabilities can be computed using Bayes' Theorem

$$p(x|y) \quad = \quad \frac{p(y|x)\,p(x)}{p(y)} \quad .$$

- When a prior distribution $\pi : \mathbb{D}\mathcal{X}$ and channel matrix $C : \mathcal{X} \to\!\triangleright \mathcal{Y}$ are clear from the context, we may use a conventional "p" notation for probabilities.

The effect of a channel on the adversary's knowledge

- First, we can define the joint distribution

$$p_{XY}(x,y) \;:=\; \pi_x \mathsf{C}_{x,y} \quad .$$

- Then by marginalization we obtain random variables X and Y s.t.

$$p_X(x) \;=\; \sum_{y:\,\mathcal{Y}} p_{XY}(x,y) \qquad \text{and} \qquad p_Y(y) \;=\; \sum_{x:\,\mathcal{X}} p_{XY}(x,y) \quad .$$

- Now we can compute the conditional probabilities (if $p(x)$, $p(y)$ are nonzero):

$$p(y|x) \;=\; \frac{p(x,y)}{p(x)} \qquad \text{and} \qquad p(x|y) \;=\; \frac{p(x,y)}{p(y)} \quad .$$

- We have that p_{XY} is the unique joint distribution that recovers π and C:

$$p(x) = \pi_x \qquad \text{and} \qquad p(y|x) = \mathsf{C}_{x,y} \;.$$

The effect of a channel on the adversary's knowledge

- For a given y, the conditional probabilities $p(x|y)$ for each x in $\mathcal{X}$ form the **posterior distribution** $p_{X|y}$.

  This represents the posterior knowledge that the adversary has about input X after observing output y.

- Each posterior distribution $p_{X|y}$ for y in $\mathcal{Y}$:

  1. represents a different state of knowledge, or "world", that an adversary seeing the output of C can end up in; and
  2. the posterior occurs when the output is y, and that output y itself occurs with probability $p(y)$.

The effect of a channel on the adversary's knowledge

- Example 5 Let's compute the posterior distributions and their probabilities for channel

C	y_1	y_2	y_3	y_4
x_1	$1/2$	$1/2$	0	0
x_2	0	$1/4$	$1/2$	$1/4$
x_3	$1/2$	$1/3$	$1/6$	0

, under the prior $\pi = (1/4, 1/2, 1/4)$.

We can write the joint distribution p_{XY} that C defines as a joint matrix J where $\mathsf{J}_{x,y} = p(x,y) = \pi_x \mathsf{C}_{x,y}$.

J	y_1	y_2	y_3	y_4
x_1	$1/8$	$1/8$	0	0
x_2	0	$1/8$	$1/4$	$1/8$
x_3	$1/8$	$1/12$	$1/24$	0

. Note that J is computed by multiplying row x of C by the prior probability π_x.

The effect of a channel on the adversary's knowledge

- Example 5 (Continued)

The marginal distribution p_Y can be found by summing the columns of J, since

$$p_Y(y) = \sum_{x:\mathcal{X}} p(x, y) \ .$$

So we get

$$p_Y = (1/4, 1/3, 7/24, 1/8) \ .$$

Each value of Y gives a posterior distribution on X, by Bayesian updating.

Since

$$p(x|y) = p(x,y)/p(y) \ ,$$

the posterior distributions can be calculated by normalizing the columns of J.

The effect of a channel on the adversary's knowledge

- Example 5 (Continued)

Hence we get the posteriors:

	$p_{X\|y_1}$	$p_{X\|y_2}$	$p_{X\|y_3}$	$p_{X\|y_4}$
x_1	$1/2$	$3/8$	0	0
x_2	0	$3/8$	$6/7$	1
x_3	$1/2$	$1/4$	$1/7$	0

.

Note that:

- The posterior that appears to be the most favorable to the adversary is

$$p_{X|y_4} \quad \text{(it's a point distribution).}$$

But it only occurs with probability $p(y_4) = 1/8$.

- The posterior that appears to be the least favorable to the adversary is:

$$p_{X|y_2} \quad \text{(it's fairly uniform).}$$

But it is the most likely to occur, with probability $p(y_2) = 1/3$. ◁

The effect of a channel on the adversary's knowledge

- The following is a fundamental result relating joints and priors.

Theorem (4.3)

For any $\pi : \mathbb{D}\mathcal{X}$ and channel matrix $\mathsf{C} : \mathcal{X} \to\!\triangleright \mathcal{Y}$, the average (i.e. expected value) of the posterior distributions $p_{X|y}$, weighted by their probabilities $p(y)$, is equal to the prior π:

$$\sum_{y \in \mathcal{Y}} p(y) p_{X|y} \quad = \quad \pi \quad ,$$

where within the summation we are multiplying a distribution (on the right) by a scalar (on the left).

It is understood pointwise: thus more explicitly we could write

$$\sum_{y \in \mathcal{Y}} p(y)\, p_{X|y}(x) = \pi_x .$$

The effect of a channel on the adversary's knowledge

- **Proof.** Weighting a posterior distribution $p_{X|y}$ with its probability $p(y)$ recovers the corresponding column of the joint matrix J.

  Hence summing those weighted posterior distributions gives the marginal distribution p_X, which is equal to the prior π. □

From joint distributions to hyper-distributions

From joint distributions to hyper-distributions

- Now we'll move from considering only the joint distribution p_{XY}.
- The first reason is that the particular output set $\mathcal{Y}$ is an inessential detail of a channel matrix C, as far as leakage is concerned.

  For an information-theoretic adversary, all deductions depend only on the correlation between input values and output values, and not on the particular output values themselves.

From joint distributions to hyper-distributions

- Example 6 Consider a channel C taking as input a user's medical record X, assumed to contain both:
  - non-sensitive data (e.g. the state where the patient resides, say Florida), and
  - sensitive data (e.g. the patient's diagnoses).

  Suppose that C is malicious and outputs:
  - *"diabetes"* if the patient has been diagnosed with diabetes, and
  - *"no diabetes"* if not.

  Remembering that the adversary is assumed to know the channel matrix C, it makes no difference in terms of leakage if:
  - the output *"diabetes"* were renamed to *"Florida"*, and
  - the output *"no diabetes"* were renamed to *"FL"*,

  according to whether or not he has diabetes.

  All that matters is the correlation between outputs and secret values, from which the adversary can infer sensitive information. ◁

From joint distributions to hyper-distributions

- A second issue with using joint distributions consist in two complications that can arise in going from a prior π and channel matrix C to the posterior distributions $p_{X|y}$ and their ($\mathcal{Y}$-marginal) probabilities $p(y)$.
  - An output value y can be impossible, in that $p(y) = 0$.

    In that case, column y of the joint matrix is all zeroes and can't be normalized; so it doesn't contribute a posterior distribution at all.
  - Two output values y and y' can give rise to the same posterior distribution, meaning that the adversary gets no benefit from distinguishing between them.

From joint distributions to hyper-distributions

- **Example 7** Suppose that prior π is $(1/3, 1/3, 0, 1/3)$ and that C is

C	y_1	y_2	y_3	y_4
x_1	$1/2$	$1/6$	$1/3$	0
x_2	0	$1/3$	$2/3$	0
x_3	0	$1/2$	0	$1/2$
x_4	$1/4$	$1/4$	$1/2$	0

.

Here we find that the joint matrix and marginal distribution on Y are

J	y_1	y_2	y_3	y_4
x_1	$1/6$	$1/18$	$1/9$	0
x_2	0	$1/9$	$2/9$	0
x_3	0	0	0	0
x_4	$1/12$	$1/12$	$1/6$	0

and $p_Y = (1/4, 1/4, 1/2, 0)$,

and therefore $p_{X|y_4}$ is undefined.

From joint distributions to hyper-distributions

- Example 7 (Continued)

Since $p(y_4) = 0$, we can drop that column, and get posterior distributions

	$p_{X\mid y_1}$	$p_{X\mid y_2}$	$p_{X\mid y_3}$
x_1	$2/3$	$2/9$	$2/9$
x_2	0	$4/9$	$4/9$
x_3	0	0	0
x_4	$1/3$	$1/3$	$1/3$

.

But now note that

$$p_{X\mid y_2} = p_{X\mid y_3},$$

and hence there are only two possible adversary "worlds", as it makes no difference to the adversary whether the output is y_2 or y_3.

Moreover, the probability of the "world" $(2/9, 4/9, 0, 1/3)$ is actually $p(y_2) + p(y_3) = 1/4 + 1/2 = 3/4$.

From joint distributions to hyper-distributions

- Example 7 (Continued)

The above suggests that we could reduce the matrix again, this time merging columns y_2 and y_3 together, obtaining

	$p_{X\|y_1}$	$p_{X\|\text{“}y_2 \text{ or } y_3\text{”}}$
x_1	$2/3$	$2/9$
x_2	0	$4/9$
x_3	0	0
x_4	$1/3$	$1/3$

.

Later, we will see the same “reduction” process applied to channel matrices, for similar reasons. ◁

From joint distributions to hyper-distributions

- The above example illustrates that the effect of channel C on prior π can be captured simply as a distribution on posterior distributions.

  We call that a **hyper-distribution**, denote it by

  $$[\pi \triangleright C] \ ,$$

  and pronounce it *"π through C"*.

- Example 8 The hyper-distribution for our example above is

$[\pi \triangleright C]$	$1/4$	$3/4$
x_1	$2/3$	$2/9$
x_2	0	$4/9$
x_3	0	0
x_4	$1/3$	$1/3$

  Note that the information in a hyper-distribution is all that we need to know the adversary's a posterior knowledge; the labels of Y are irrelevant. $\triangleleft$

From joint distributions to hyper-distributions

- The abstraction afforded by hyper-distributions is particularly important when we want to *compare* which channel is more secure.

- **Example 9** Given $\mathcal{X} = \{x_1, x_2, x_3\}$, consider channel matrices C and D:

C	y_1	y_2	y_3
x_1	1	0	0
x_2	$1/4$	$1/2$	$1/4$
x_3	$1/2$	$1/3$	$1/6$

D	z_1	z_2	z_3
x_1	$2/5$	0	$3/5$
x_2	$1/10$	$3/4$	$3/20$
x_3	$1/5$	$1/2$	$3/10$

.

Although C and D look different, they are identical wrt. the leakage of X they cause.

Both map an arbitrary prior $\pi = (p_1, p_2, p_3)$ to the very same hyper-distribution on the side.

In terms of leakage, the channels are the same!

	$\frac{4p_1+p_2+2p_3}{4}$	$\frac{3p_2+2p_3}{4}$
x_1	$\frac{4p_1}{4p_1+p_2+2p_3}$	0
x_2	$\frac{p_2}{4p_1+p_2+2p_3}$	$\frac{3p_2}{3p_2+2p_3}$
x_3	$\frac{2p_3}{4p_1+p_2+2p_3}$	$\frac{2p_3}{3p_2+2p_3}$

◁

From joint distributions to hyper-distributions

- We are now ready to formalize hyper-distributions.

Definition (4.5)

If $\mathcal{X}$ is a finite set (of possible secret values), a **hyper-distribution** (or just a **hyper**) Δ is a distribution on distributions on $\mathcal{X}$, so that Δ has type $\mathbb{D}(\mathbb{D}\mathcal{X})$, which we abbreviate to $\mathbb{D}^2\mathcal{X}$.

We recall that the support $\lceil\Delta\rceil$ of Δ is the set of distributions to which Δ gives positive probability (i.e. $\lceil\Delta\rceil = \{\delta\colon \mathbb{D}\mathcal{X} \mid \Delta_\delta > 0\}$) and we assume that set to be finite; they are the set of possible "worlds" under Δ and we call them the **inner distributions**, or just the **inners**, of Δ.

We call the distribution on the inners themselves as the **outer distribution** of Δ.

As mentioned above, the hyper-distribution that results from a channel matrix C on prior π is written $[\pi \triangleright C]$, and pronounced "π through C".

When we want to have names for the inners and their outer probabilities, we write $[\pi \triangleright C] = \sum_i a_i[\delta^i]$ to indicate that $[\pi \triangleright C]$ has inners δ^i and outer probabilities a_i.

From joint distributions to hyper-distributions

- **Example 10** Consider again the channel

C	y_1	y_2	y_3	y_4
x_1	$1/2$	$1/2$	0	0
x_2	0	$1/4$	$1/2$	$1/4$
x_3	$1/2$	$1/3$	$1/6$	0

, under the prior $\pi = (1/4, 1/2, 1/4)$.

In a previous example we found that the corresponding joint distribution is

J	y_1	y_2	y_3	y_4
x_1	$1/8$	$1/8$	0	0
x_2	0	$1/8$	$1/4$	$1/8$
x_3	$1/8$	$1/12$	$1/24$	0

, and the marginal on Y is $p_Y = (1/4, 1/3, 7/24, 1/8)$.

From joint distributions to hyper-distributions

- Example 10 (Continued)

Hence, the corresponding hyper-distribution is

$[\pi \triangleright C]$	$1/4$	$1/3$	$7/24$	$1/8$
x_1	$1/2$	$3/8$	0	0
x_2	0	$3/8$	$6/7$	1
x_3	$1/2$	$1/4$	$1/7$	0

.

Here the outer distribution is

$$(1/4, 1/3, 7/24, 1/8)$$

on the inner distributions

- $(1/2, 0, 1/2)$,
- $(3/8, 3/8, 1/4)$,
- $(0, 6/7, 1/7)$, and
- $(0, 1, 0)$, respectively.

From joint distributions to hyper-distributions

- Notation. Given a prior π and a channel matrix C:
  - $\pi \triangleright C$ denotes the joint matrix J resulting from pushing π through C; and
  - $[\pi \triangleright C]$ denotes the hyper distribution resulting from pushing π through C.
- The brackets $[-]$ on their own act as the function that:
  - takes a joint matrix of type $\mathbb{D}(\mathcal{X} \times \mathcal{Y})$, and
  - maps it to the corresponding hyper-distribution of type $\mathbb{D}^2\mathcal{X}$.
- Fundamental principle:

The information-theoretic essence of a channel matrix C is a mapping from priors π to hyper-distributions $[\pi \triangleright C]$.

Abstract channels

Abstract channels

- As we have seen, the information-theoretic leakage of a channel matrix C is determined by its mapping from prior distributions to hyper-distributions.

  We call that mapping the abstract channel denoted by C.

Definition (4.7)

The **abstract channel** C denoted by channel matrix C is the mapping of type $\mathbb{D}\mathcal{X} \to \mathbb{D}^2\mathcal{X}$ that C gives, namely $\pi \mapsto [\pi \triangleright \mathsf{C}]$.

We use semantic brackets for that denotation, writing $C = [\![\mathsf{C}]\!]$.

- Notation:
  - C denotes a **concrete channel** of type $\mathsf{C}\colon \mathcal{X} \to\!\triangleright \mathcal{Y}$ (typically represented by a channel matrix);
  - C denotes an **abstract channel** of type $C\colon \mathbb{D}\mathcal{X} \to \mathbb{D}^2\mathcal{X}$.

Abstract channels

- Abstract channels are of central importance in QIF.
  - Channel matrices contain detail that is extraneous to their fundamental meaning (labels in $\mathcal{Y}$, order of rows, etc.)

    Abstract channels keep only relevant information wrt. information leakage.
  - Abstract channel facilitate the mathematical approach to other results we'll see in this course (including on composition and on refinement of channels).
- It's important to note that when analyzing matters other than information leakage (e.g., system functionality), the use of concrete channels may be more appropriate than the use of abstract channels.

Abstract channels

- We now present important properties of abstract channels.
- First, Theorem 4.3 can be extended also to them.

Corollary (4.8)

For any abstract channel C and prior π, the average of the inner distributions of $[\pi \triangleright C]$, weighted by their outer probabilities, is equal to the prior π. That is,

$$\pi \quad = \quad \sum_i a_i \delta^i \ , \qquad where \qquad [\pi \triangleright C] \quad = \quad \sum_i a_i [\delta^i] \quad .$$

Proof. The only difference from 4.3 is that $[\pi \triangleright C]$ merges any posterior distributions that are equal, summing their outer probabilities — and that has no effect on the expectation. □

Abstract channels

- Note that the expected value of hyper-distribution $[\pi \triangleright C]$ (which is a distribution on distributions) is the prior distribution π.

  Writing $\mu\Delta$ for the expected value of a distribution Δ, so that

$$\mu\Delta = \sum_{\delta} \Delta_\delta \times \delta,$$

  the corollary above can be written as

$$\mu[\pi \triangleright C] = \pi \ .$$

Abstract channels

- We now find a concrete representation of abstract channels.

  Recall that channel matrices contain extraneous structure:

  1. labels on columns,
  2. columns that are all zero, representing outputs that can never occur, and
  3. similar columns, by which we mean columns that are scalar multiples of each other and therefore yield the same posterior distributions on all priors.

- By eliminating this extraneous structure to get a well defined reduced matrix.

Definition (4.9)

The **reduced matrix** C^r of a channel matrix C is formed by deleting output labels and all-zero columns, adding similar columns together, and finally ordering the resulting columns lexicographically.

The reduced matrix is analogous to a **normal form**, also a syntactic concept.

Abstract channels

- The following results shows that the reduction process preserves meaning.

Theorem (4.10)

Any channel matrix C *denotes the same abstract channel as its reduced matrix* C^r. *That is,*

$$[\![C]\!] = [\![C^r]\!].$$

Proof. Output labels, all-zero columns, and column ordering all have no effect on the hyper-distribution.

And similar columns each contribute weight to the same posterior distribution; hence merging them leaves the hyper-distribution unchanged. □

Abstract channels

- The following results shows that the reduction process preserves meaning.

Corollary (4.11)

Two channel matrices C *and* D *denote the same abstract channel iff their reduced matrices are equal: that is,* $[\![C]\!]=[\![D]\!]$ *iff* $C^r=D^r$.

Proof. From 4.10 we have $[\![C]\!] = [\![D]\!]$ iff $[\![C^r]\!] = [\![D^r]\!]$.

And it is easy to see that $[\![C^r]\!] = [\![D^r]\!]$ iff $C^r = D^r$: the backward implication is immediate, and the forward implication follows because distinct reduced matrices cannot map the uniform prior (for instance) to the same hyper-distribution. □

- The result is that a reduced matrix (like a normal form) serves as a canonical syntactic representation of an abstract channel.

Abstract channels

- **Example 11** Consider again the channel matrices C and D:

C	y_1	y_2	y_3
x_1	1	0	0
x_2	$1/4$	$1/2$	$1/4$
x_3	$1/2$	$1/3$	$1/6$

D	z_1	z_2	z_3
x_1	$2/5$	0	$3/5$
x_2	$1/10$	$3/4$	$3/20$
x_3	$1/5$	$1/2$	$3/10$

.

Let us find their corresponding reduced matrices.

- In C the columns y_2 and y_3 are similar: column y_2 is 2 times column y_3.
- In D the columns z_1 and z_3 are similar: column z_1 is $2/3$ times column z_3.

Merging similar columns, we find that C and D have the *same* reduced matrix:

$$C^r \quad = \quad D^r \quad = \quad \begin{array}{|c|cc|} \hline x_1 & 1 & 0 \\ x_2 & 1/4 & 3/4 \\ x_3 & 1/2 & 1/2 \\ \hline \end{array} \quad .$$

That shows that C and D denote the same abstract channel. ◁

Abstract channels

- Now we extend the concept *determinism* from channel matrices to abstract channels.

Definition (4.13)

We say that an abstract channel C is **deterministic** just when, for every π, the inners of $[\pi \triangleright C]$ have pairwise-disjoint supports.

(The corresponding concept for concrete channels is that their reductions contain only zeroes and ones.)

- Note that a channel matrix C is essentially deterministic iff its corresponding abstract channel C *is* deterministic according to Definition 4.13.

Abstract channels

- **Example 12** Consider channel matrix C below.

C	y_1	y_2	y_3	y_4
x_1	$2/3$	0	$1/3$	0
x_2	0	$3/4$	0	$1/4$
x_3	$2/3$	0	$1/3$	0

We can check that under any prior π the corresponding hyper $[\pi \triangleright \mathsf{C}]$ presents only inners with pairwise-disjoint supports.

Indeed, C's reduced matrix is

$$\mathsf{C}^r = \begin{array}{|c|cc|} \hline x_1 & 1 & 0 \\ x_2 & 0 & 1 \\ x_3 & 1 & 0 \\ \hline \end{array} \quad ,$$

and the hyper $[\pi \triangleright \mathsf{C}^r]$ for any prior $\pi = (p_1, p_2, p_3)$ is

$[\pi \triangleright \mathsf{C}^r]$	$p_1 + p_3$	p_2
x_1	$p_1/(p_1+p_3)$	0
x_2	0	1
x_3	$p_3/(p_1+p_3)$	0

.

◁

Abstract channels

- There are two extreme channels that deserve consideration:
  - The abstract channel that leaks nothing is the mapping

$$\pi \mapsto [\pi] \ .$$

    The corresponding reduced channel matrix is a single column of 1's.

    We call that channel $\mathbb{1}$, as it preserves secrecy.

  - The abstract channel that leaks everything is the mapping

$$\pi \mapsto \sum_x \pi_x [[x]] \ .$$

    The corresponding reduced channel matrix is the identity matrix I.

    We call that channel $\mathbb{0}$, as it annihilates secrecy.

Abstract channels

- **Example 13** Consider channel matrices C and D below.

C	y_1	y_2	y_3
x_1	$1/2$	$1/3$	$1/6$
x_2	$1/2$	$1/3$	$1/6$
x_3	$1/2$	$1/3$	$1/6$

D	y_1	y_2	y_3	y_4	y_5
x_1	0	$1/3$	0	0	$2/3$
x_2	$3/4$	0	$1/4$	0	0
x_3	0	0	0	1	0

Their reduced matrices C^r and D^r are:

$$\mathsf{C}^r = \begin{array}{|c|c|}\hline x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \hline \end{array} \qquad \mathsf{D}^r = \begin{array}{|c|c|c|c|}\hline x_1 & 1 & 0 & 0 \\ x_2 & 0 & 1 & 0 \\ x_3 & 0 & 0 & 1 \\ \hline \end{array}$$

Under any prior $\pi = (p_1, p_2, p_3)$ the corresponding hypers $[\pi \triangleright \mathsf{C}^r]$ and $[\pi \triangleright \mathsf{D}^r]$ are:

$[\pi \triangleright \mathsf{C}^r]$	1
x_1	p_1
x_2	p_2
x_3	p_3

$[\pi \triangleright \mathsf{D}^r]$	p_1	p_2	p_3
x_1	1	0	0
x_2	0	1	0
x_3	0	0	1

Hence we have that:

- $[\![\mathsf{C}]\!] = \mathbb{1}$ is the channel that preserves privacy.
- $[\![\mathsf{D}]\!] = \mathbb{O}$ is the channel that annihilates privacy. ◁

Abstract channels

- A traditional name for the "no leakage" property of $\mathbb{1}$ is noninterference.

  On channel matrices this is the property that the output is completely independent of the input.

Definition (4.14)

A channel matrix C satisfies **noninterference** just when its rows are all the same, which means that its reduction C^r consists of a single column of 1's.

Equivalently, its abstract channel is the mapping $\pi \mapsto [\pi]$.

(Channel $\mathbb{1}$ is hence the canonical noninterfering channel.)

Appendix: More on abstract channels

More on abstract channels

- As we have seen (Def. 4.7), an abstract channel C is a mapping

$$\mathbb{D}\mathcal{X} \to \mathbb{D}^2\mathcal{X}$$

  denoted by some channel matrix C, so that $C = [\![\mathsf{C}]\!]$.

  Here we explore such mappings a bit further.

- We already noted two special properties of abstract channels:

  1. Any abstract channel C always produces hyper-distributions with only *finitely many inners*, because any channel matrix has only finitely many columns.
  2. For any abstract channel C and prior π, the expected value of the hyper-distribution $[\pi \triangleright C]$ is always the prior π.

- We might wonder whether those two properties are enough to characterize the abstract channels that are denoted by channel matrices.

  The following counterexample shows that they are not.

More on abstract channels

- **Example 14** Let $\mathcal{X} = \{x_1, x_2\}$ and let

$$C : \mathbb{D}\mathcal{X} \to \mathbb{D}^2\mathcal{X}$$

map prior

$$\pi = (a, 1-a), \quad \text{for } 0 \leq a \leq 1,$$

to the hyper-distribution

$[\pi \triangleright C]$	a^2	$1 - a^2$
x_1	1	$\frac{a}{1+a}$
x_2	0	$\frac{1}{1+a}$

.

Note that this hyper-distribution indeed has finitely many inners, and its expectation is indeed the prior π.

(Also notice that if a=0 or a=1 then the hyper-distribution actually has just one inner.)

More on abstract channels

- **Example 14** (Continued)

But it is easy to see that C is not the denotation of any channel matrix.

For if $0 < a < 1$, then we can work backwards to recover a unique reduced channel matrix C such that $[\pi \triangleright \mathsf{C}] = [\pi \triangleright C\,]$.

First we multiply each inner by its outer probability to recover the joint matrix

$$\mathsf{J} \quad = \quad \begin{array}{|c|cc|} \hline x_1 & a^2 & a(1-a) \\ x_2 & 0 & 1-a \\ \hline \end{array} \quad ,$$

and then by normalizing the rows we obtain the reduced channel matrix

$$\mathsf{C} \quad = \quad \begin{array}{|c|cc|} \hline x_1 & a & 1-a \\ x_2 & 0 & 1 \\ \hline \end{array} \quad .$$

But now we see that no single C will do the job, since this C *varies* with the prior $\pi = (a, 1-a)$. ◁

More on abstract channels

- The same reasoning used in the Example above lets us prove the following theorem, which shows that the abstract channels denoted by channel matrices are very tightly constrained indeed.

Theorem

Let Δ be a hyper-distribution with finitely many inners. Then there exists an abstract channel C and a unique prior π such that $[\pi \triangleright C] = \Delta$. And if π is full support, then also C is unique.

Proof. In the book! $\square$

Appendix

A first look at channel compositions: Convex combinations of channels

A first look at channel compositions

- So far we have viewed channels as being monolithic.

  But, from the perspective of modularity, it is valuable to consider channels that are composed in some way from smaller channels.

- Next we briefly consider some channel compositions (which we'll cover with more detail later in this course).

A first look at channel compositions: Convex combinations of channels

- Recall that a convex combination is the averaging of a set of elements using coefficients c_i that are non-negative and sum to one.

- Example 15 The convex combination of distributions δ^i is a new distribution

$$\delta = \sum_i c_i \delta^i.$$

◁

- Example 16 Hypers, in $\mathbb{D}^2\mathcal{X}$, are thus distributions over elements in $\mathbb{D}\mathcal{X}$ — and hence they too support convex combination. For example

$$\Delta = \sum_i c_i \Delta^i$$

is a new hyper, assigning to each inner δ the new outer probability

$$\Delta_\delta = \sum_i c_i \Delta^i_\delta.$$

◁

A first look at channel compositions: Convex combinations of channels

- Note that only the outer probabilities are affected by that operation.

  The inners of Δ are simply the union of the inners of all Δ^i's (excluding the degenerate case where some c_i is 0).

  The fact that hypers can be combined that way is the reason we write

$$[\pi \triangleright C] = \sum_i a_i [\delta^i]$$

  to denote that $[\pi \triangleright C]$ has inners δ^i and outer probabilities a_i.

  Each $[\delta^i]$ is a point hyper, with a single inner δ^i, hence their convex combination with the outer probabilities a_i forms the hyper $[\pi \triangleright C]$.

A first look at channel compositions: Convex combinations of channels

- As a consequence, abstract channels can be also constructed by convex combination of their produced hypers:

$$C := \sum_i c_i C^i$$

is the abstract channel such that

$$[\pi \triangleright C] = \sum_i c_i [\pi \triangleright C^i].$$

Here C can be viewed as a channel that probabilistically chooses some C^i, revealing which one was chosen, and then behaves like C^i.

That operation is called external choice and we'll study it further in this course.

A first look at channel compositions: Convex combinations of channels

- Example 17 (**External choice**) Imagine a scenario where the adversary probabilistically gets a report from one of two spies:
  - Boris (with probability p), or
  - Natasha (with probability $1-p$).

  The reports are handwritten, so that the adversary can tell who was the author.

  Hence, writing B and N for Boris's and Natasha's abstract channels, the hyper-distribution that the adversary gets is the convex combination

$$p[\pi \triangleright B] + (1-p)[\pi \triangleright N].$$

◁

A first look at channel compositions: Convex combinations of channels

- External choice can also be done on the channel matrices B and N, provided that we rename output labels as necessary to make the output sets of B and N disjoint (corresponding to distinguishable handwriting).

  Then:

  - the set of columns of the resulting channel matrix

    $$p\mathrm{B} + (1-p)\mathrm{N}$$

    is simply the union of the columns of B and N; and
  - the entries in columns from B are scaled with p while those from N are scaled with $1-p$.

A first look at channel compositions: Convex combinations of channels

- Example 18 (**Internal choice**) Now suppose that Boris and Natasha write their reports on a typewriter, making them indistinguishable, and moreover that they use the very same set $\mathcal{Y}$ of possible outputs.

  That's another way of combining channel matrices, again convex.

  If C^i are channel matrices from $\mathcal{X}$ to $\mathcal{Y}$, then

  $$\mathsf{C} = \sum_i c_i \mathsf{C}^i$$

  is simply the convex combination of the corresponding matrices.

  (Note how that differs from external choice on abstract channels, where each C^i could have had its own $\mathcal{Y}^i$ when formulated as a channel matrix C^i.)

  Here C can be viewed as a channel that probabilistically chooses some C^i, without revealing which one was chosen, and then behaves like C^i.

  That operation is called internal choice, and it's fundamentally different from external choice.

A first look at channel compositions: Convex combinations of channels

- Example 19 The effect of internal choice depends critically on the "accidental" detail of how X is correlated with the values in Y for each C^i.

  In particular, permuting the output labels of some C^i has no effect on the abstract channel that it denotes, but it could greatly affect the resulting internal choice.

A first look at channel compositions: Convex combinations of channels

- Example 19 (Continued)

To see that, suppose Boris and Natasha use the following channel matrices:

B	y_1	y_2
x_1	1	0
x_2	0	1

N	y_1	y_2
x_1	0	1
x_2	1	0

.

Although each of those channel matrices leaks X completely, both denoting the leak-everything channel $\mathbb{O}$, their output labels have opposite meanings.

As a result, their convex combination

$$1/2\mathsf{B} + 1/2\mathsf{N}$$

as matrices leaks nothing.

(It's the matrix with $1/2$ in every position, which reduces to the single-column matrix of all 1's, i.e. the channel $\mathbb{1}$).

A first look at channel compositions: Convex combinations of channels

- Example 19 (Continued)

  Yet, if we take the convex combination of the corresponding abstract channels, we have $B = N$, and hence

  $$C = {}^1\!/_2 B + {}^1\!/_2 N$$

  also leaks X completely.

  Moreover, swapping the output labels y_1 and y_2 in B has no effect on B, but it causes the convex combination ${}^1\!/_2\mathrm{B} + {}^1\!/_2\mathrm{N}$ also to leak X completely.

  Note finally that internal choice cannot even be defined on abstract channels, because (since they have abstracted from the output labels) there is no way to know which inners should be combined. ◁

Appendix
A first look at channel compositions: Cascading

A first look at channel compositions

- So far we have viewed channels as being monolithic.

  But, from the perspective of modularity, it is valuable to consider channels that are composed in some way from smaller channels.

- Next we briefly consider some channel compositions (which we'll cover with more detail later in this course).

A first look at channel compositions: Cascading and the Data-Processing Inequality

- We conclude with a discussion of an important and natural way of composing channel matrices known as cascading.

Definition (4.18)

Given channel matrices C: $\mathcal{X} \rightarrow\!\triangleright \mathcal{Y}$ and D: $\mathcal{Y} \rightarrow\!\triangleright \mathcal{Z}$, the **cascade** of C and D is the channel matrix CD of type $\mathcal{X} \rightarrow\!\triangleright \mathcal{Z}$, where CD is given by ordinary matrix multiplication.

- Informally, we can see:
  1. C as specifying the correlation $p(y|x)$ between X and Y;
  2. D as specifying the correlation $p(z|y)$ between Y and Z; and
  3. the cascade CD is then the resulting correlation $p(z|x)$ between X and Z.

A first look at channel compositions: Cascading and the Data-Processing Inequality

- Given prior $\pi: \mathbb{D}\mathcal{X}$ and channel matrices C and D, the joint distribution

$$p_{XYZ}(x, y, z) \;:=\; \pi_x \mathrm{C}_{x,y} \mathrm{D}_{y,z}$$

can be proven to be the unique joint distribution satisfying:

1. p_{XYZ} recovers π, so that $p(x) = \pi_x$,
2. p_{XYZ} recovers C, so that $p(y|x) = \mathrm{C}_{x,y}$,
3. p_{XYZ} recovers D, so that $p(z|y) = \mathrm{D}_{y,z}$, and
4. $p(z|x, y)$ does not depend on x, so that $p(z|x, y) = p(z|y)$.

A first look at channel compositions: Cascading and the Data-Processing Inequality

- Our above informal statement about cascading can now be made precise:

$$p(z|x) \quad = \quad (\mathrm{CD})_{x,z} \quad ,$$

where CD is just the matrix-product between C and D.

This is because

$$p(z|x) \quad = \quad \frac{p(x,z)}{p(x)} \quad = \quad \frac{\sum_y p(x,y,z)}{p(x)} \quad = \quad \frac{\sum_y \pi_x \mathrm{C}_{x,y} \mathrm{D}_{y,z}}{\pi_x} \quad = \quad (\mathrm{CD})_{x,z} \quad .$$

- Example 20 The cascading of channel matrices $\mathrm{C}: \mathcal{X} \to \mathcal{Y}$ and $\mathrm{D}: \mathcal{Y} \to \mathcal{Z}$ below is $\mathrm{CD}: \mathcal{X} \to \mathcal{Z}$ as follows.

C	y_1	y_2	y_3
x_1	$1/2$	$1/2$	0
x_2	$1/3$	$1/6$	$1/2$

$\cdot$

D	z_1	z_2
y_1	$1/3$	$2/3$
y_2	0	1
y_3	1	0

$=$

CD	z_1	z_2
x_1	$1/6$	$5/6$
x_2	$11/18$	$7/18$

$\cdot$ $\triangleleft$

A first look at channel compositions: Cascading and the Data-Processing Inequality

- Let's now consider cascade CD from the perspective of information leakage:
  1. D can be seen as a sanitization policy that specifies how the output labels $\mathcal{Y}$ of C should be mapped to the output labels $\mathcal{Z}$ of D.
  2. The composition CD suppresses the release of Y and instead releases the "sanitized" Z, thereby (perhaps) leaking less.
- We'll see later that for any "sanitation policy" D the leakage can never be increased.

  That's known as the Data-Processing Inequality, and we'll return to it later.

  But it should be obvious that different sanitization policies could be more or less effective at limiting the leakage caused by C.