
エントロピー論の現在

エントロピー論の現在
Agenda

第一部　シャノン･エントロピーの基本的な性質について

第二部　シャノンが考えたことを振り返る

第三部　Faddeev-LeinsterのChain ルール

第四部　Baez:新しいエントロピー論の登場
-- Entropy as a Functor

第一部

シャノン・エントロピーの基本的な性質について

第一部 Agenda

シャノン・エントロピーの基本的な性質について

- 確率分布とシャノン・エントロピー
- $H(X) \geq 0$
- $H(X)=0$となる場合
- $H(X)$が最大となる場合
- エントロピーの単位としての bit

確率分布とシャノン・エントロピー

シャノンのエントロピー H(X)の定義

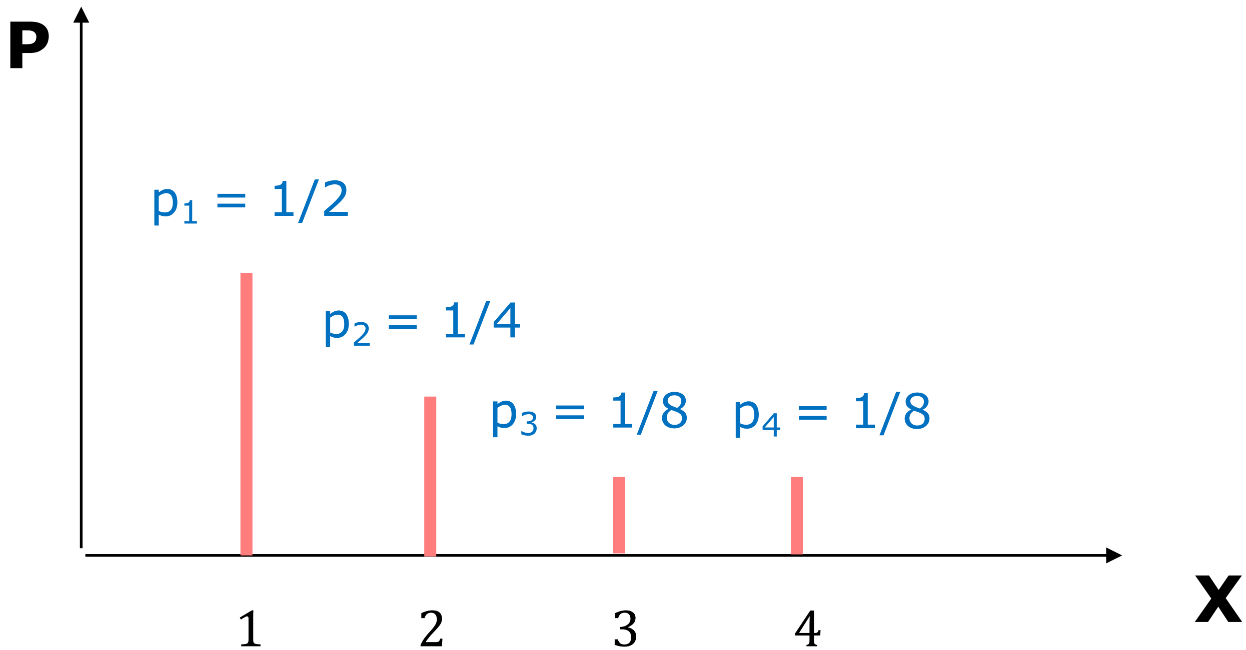
シャノンは、シャノンのエントロピーを次の式で定義した。

$$\mathrm{H(X)} = -\sum_{i} p_i \log_2 p_i$$

n個の確率変数が取る値を X={$x_1, x_2, \ldots, x_n$}、その確率分布をP=($p_1, p_2, \ldots, p_n$)とする時、H(X)はこの確率分布のエントロピーを与える。

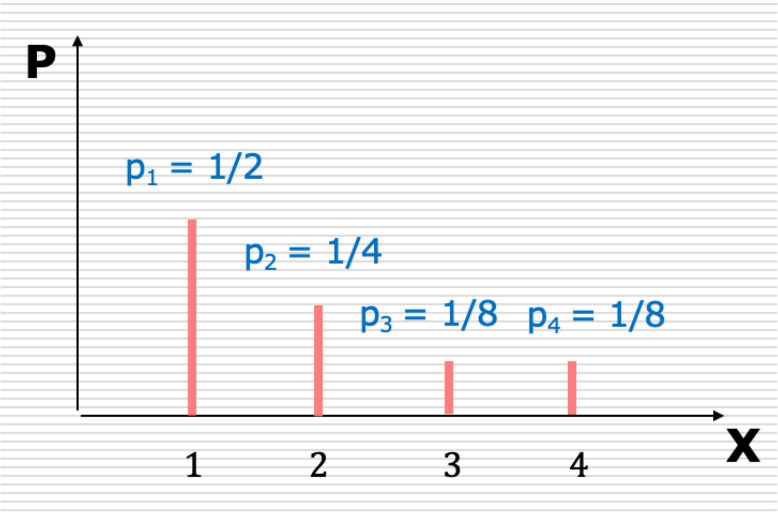
確率分布が与えられれば、エントロピーが決まる。

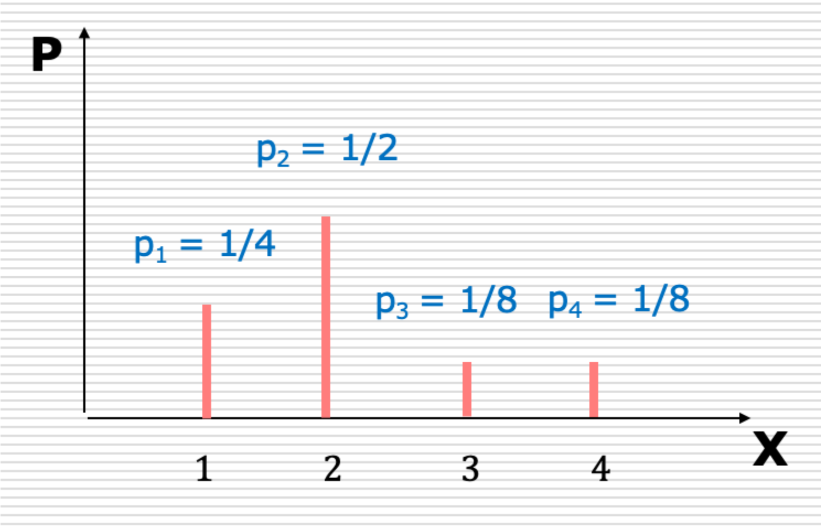
次のような確率分布が与えられた時の エントロピーは？

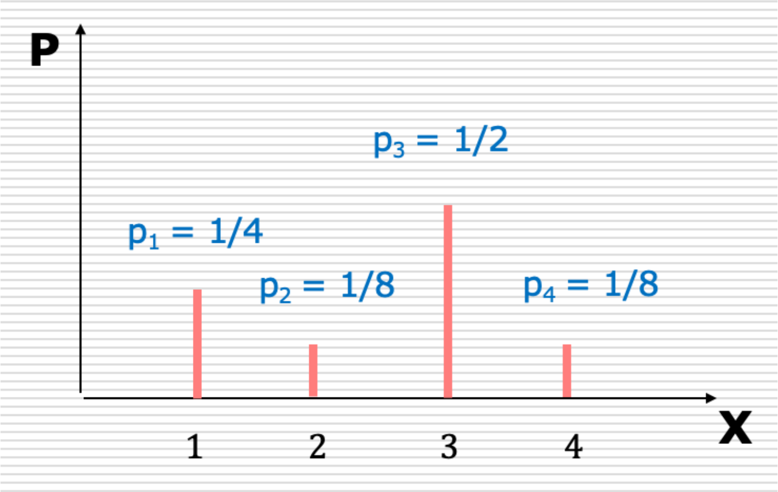


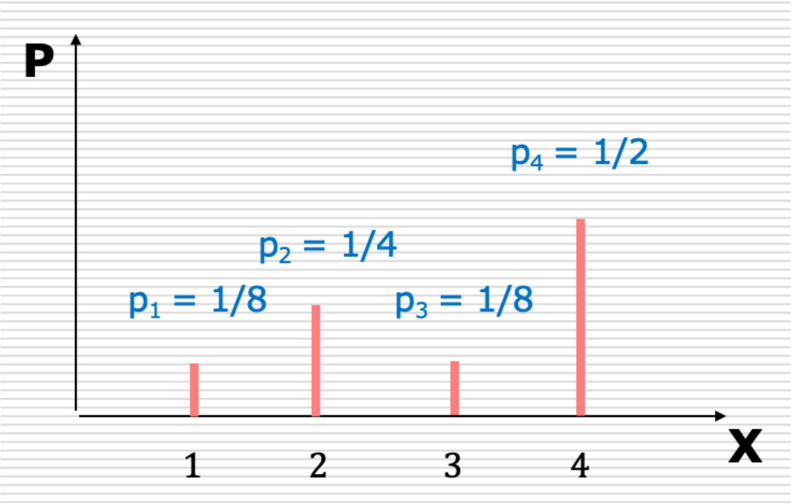
$$
\begin{aligned}
H(X) &= 1/2 \cdot \log 2 + 1/4 \cdot \log 4 + 1/8 \cdot \log 8 + 1/8 \cdot \log 8 \\
&= 1/2 \cdot \log 2 + 1/4 \cdot \log 2^2 + 1/8 \cdot \log 2^3 + 1/8 \cdot \log 2^3 \\
&= 1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/8 \cdot 3 = 1/2 + 1/2 + 2 \cdot 3/8 \\
&= 1.75
\end{aligned}
$$

次の確率分布は、分布の形は違っていても同じエントロピーを持つ

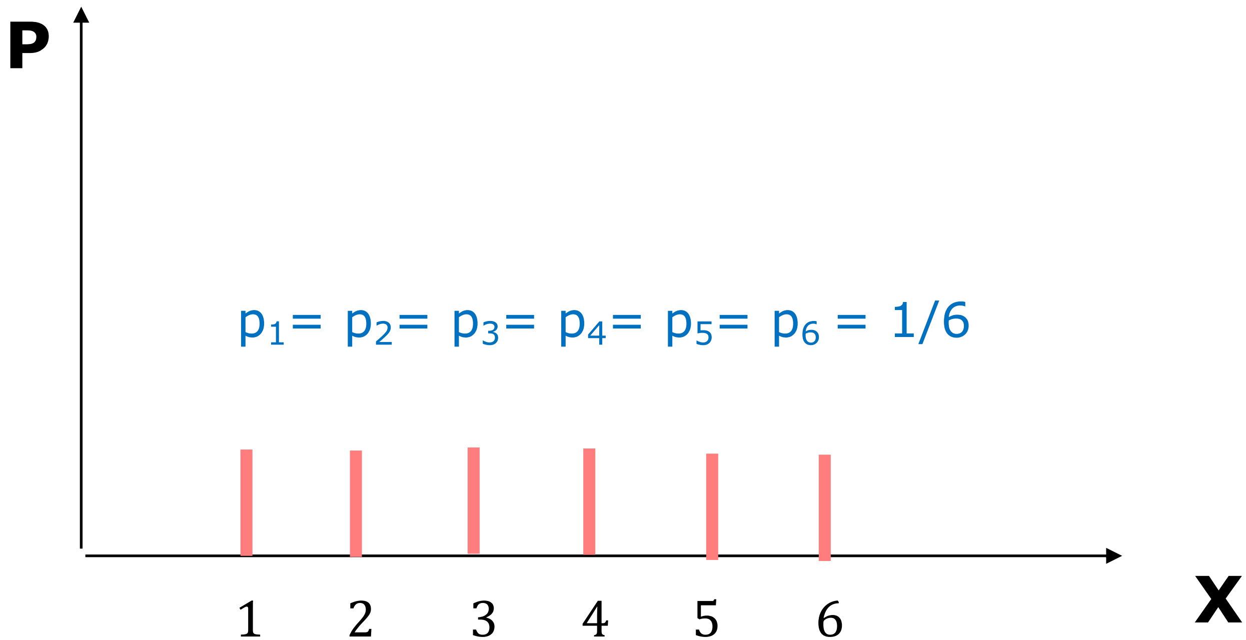








サイコロの出る目の確率分布とそのエントロピー

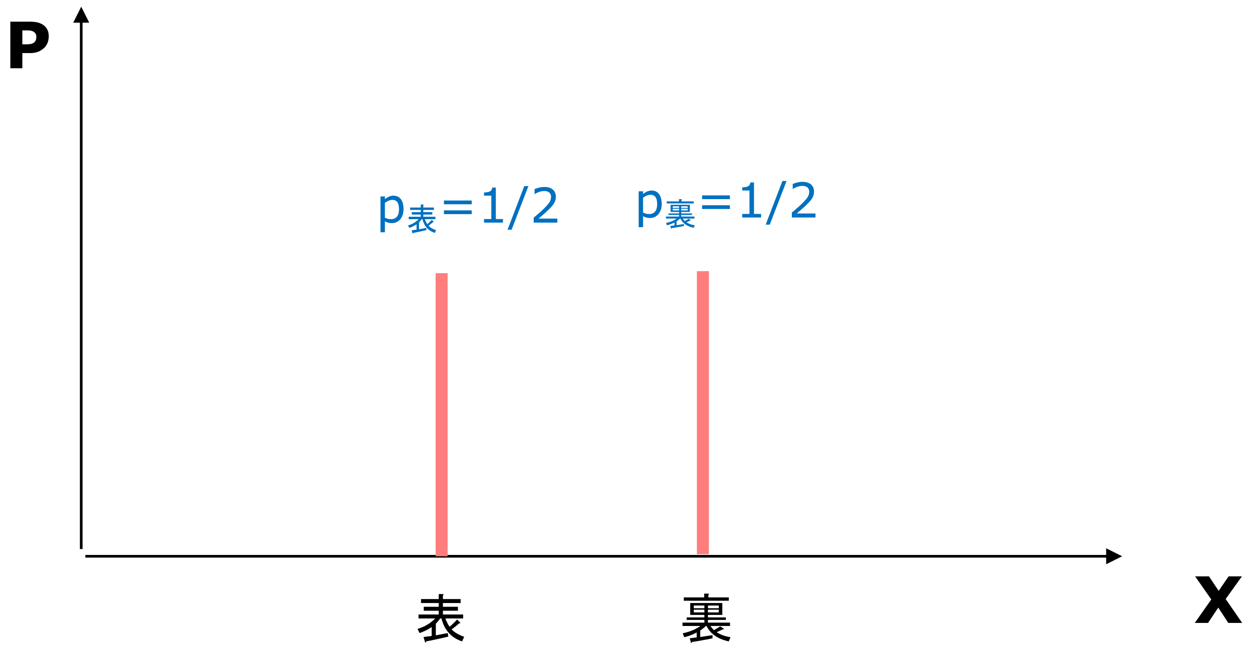


$p1=p2=p3=p4=p5=p6= 1/6$ だから

$H(X)=1/6\cdot(\log 6+\log 6+\log 6+\log 6+\log 6+\log 6)=\log 6$

$= 1+\log 3$

フェアなコイン・トスの確率分布とそのエントロピー

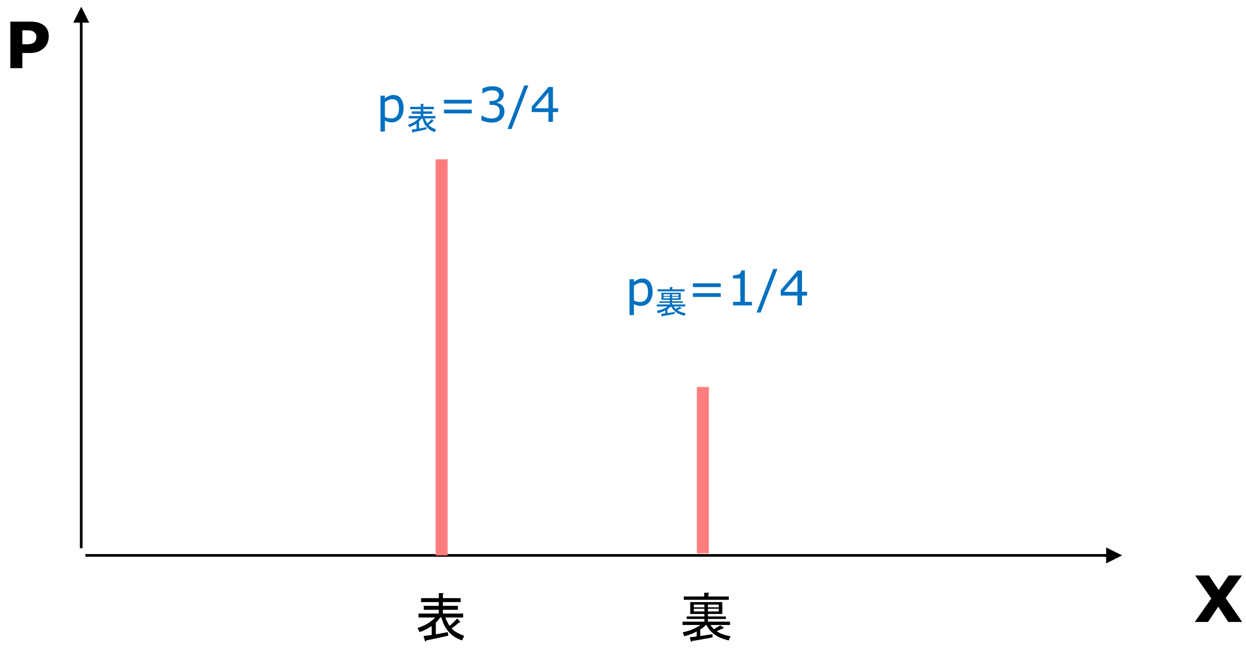


フェアなコインを投げた時、表が出る確率は1/2、裏が出る確率は1/2。

H(X) = 1/2・log 2 + 1/2・log 2 = log 2=1

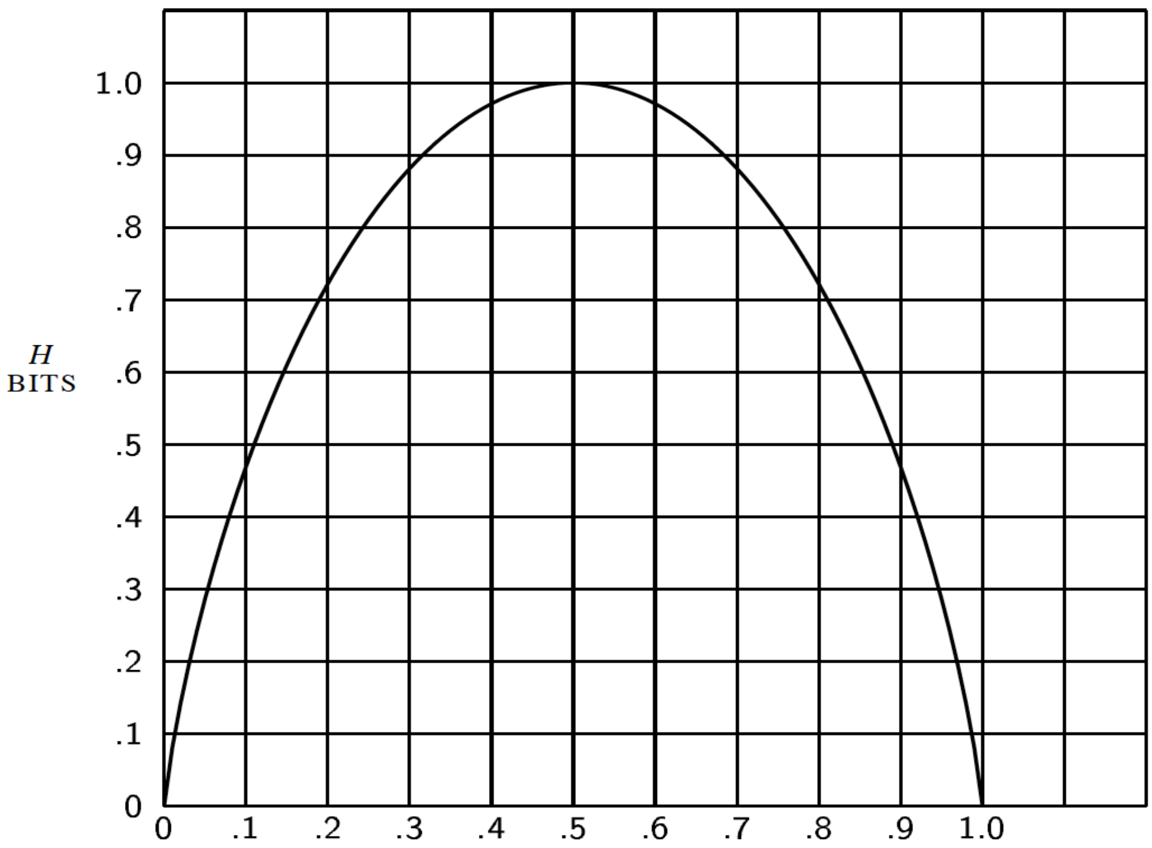
フェアなコイン・トスの確率分布のエントロピーは、1 である。

フェアでないコイン・トスの確率分布とそのエントロピー



フェアコインでないコインがある。
表が出る確率は3/4、裏が出る確率は1/4だとする。
H(X) = 3/4･log 3/4 + 1/4･log 4 = 3/4(log 3-log 4)+1/2
このフェアでないコイン･トスの確率分布のエントロピーは、3/4 log 3 – 1 。

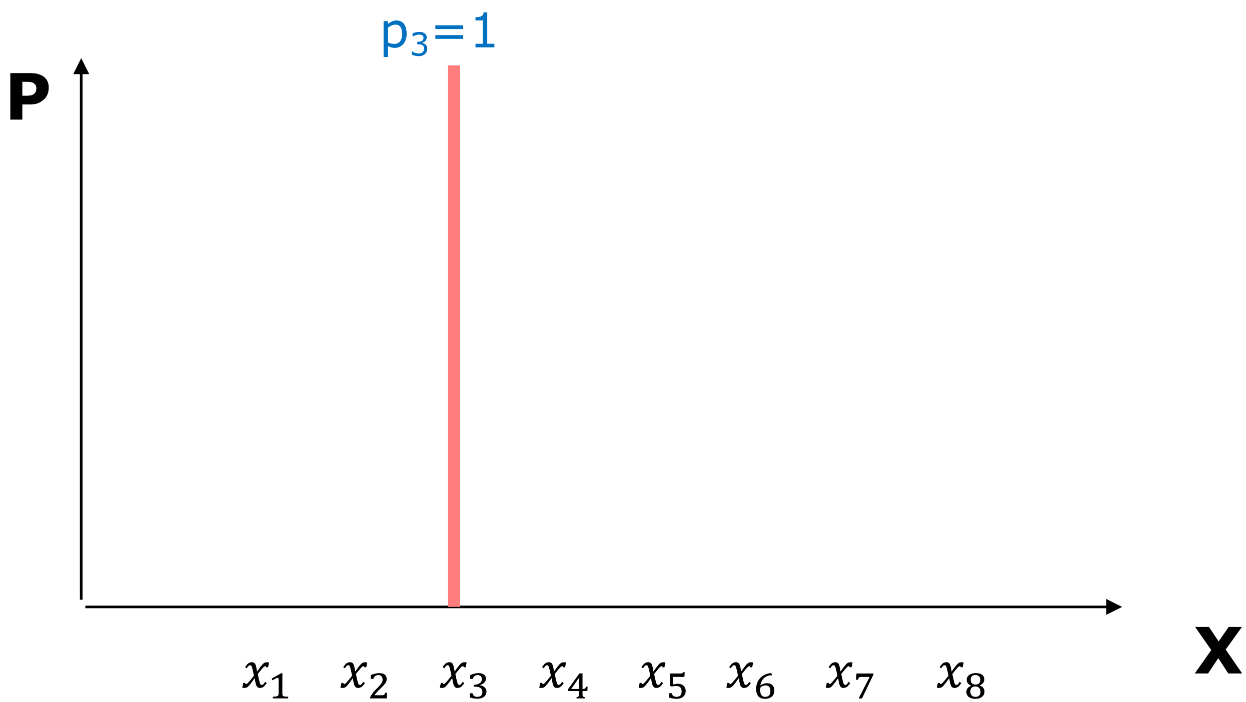
$p + q = 1$ を満たす二項の確率分布とそのエントロピーのグラフ



$p + q = 1$の時の

$H(\mathrm{p},\mathrm{q}) = -\mathrm{p}\log\mathrm{p} - \mathrm{q}\log\mathrm{q} = \mathrm{p}\log\left(\frac{1}{\mathrm{p}}\right) + \mathrm{q}\log(\frac{1}{\mathrm{q}})$ のグラフ

一つのp_iのみが1の場合のエントロピー



この例では、Xがx_3の値を取る場合のみ、p_3=1である。
この時、H(X)=1･log(1/1)=1･log1=1･0=0
（ただし、0･log(1/0）=0とした）

シャノン・エントロピーの性質について 1

$$H(X) \geq 0$$

$\mathrm{H(X)} \geq 0$

$$\mathrm{H(X)} = -\sum_i p_i \log_2 p_i$$

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

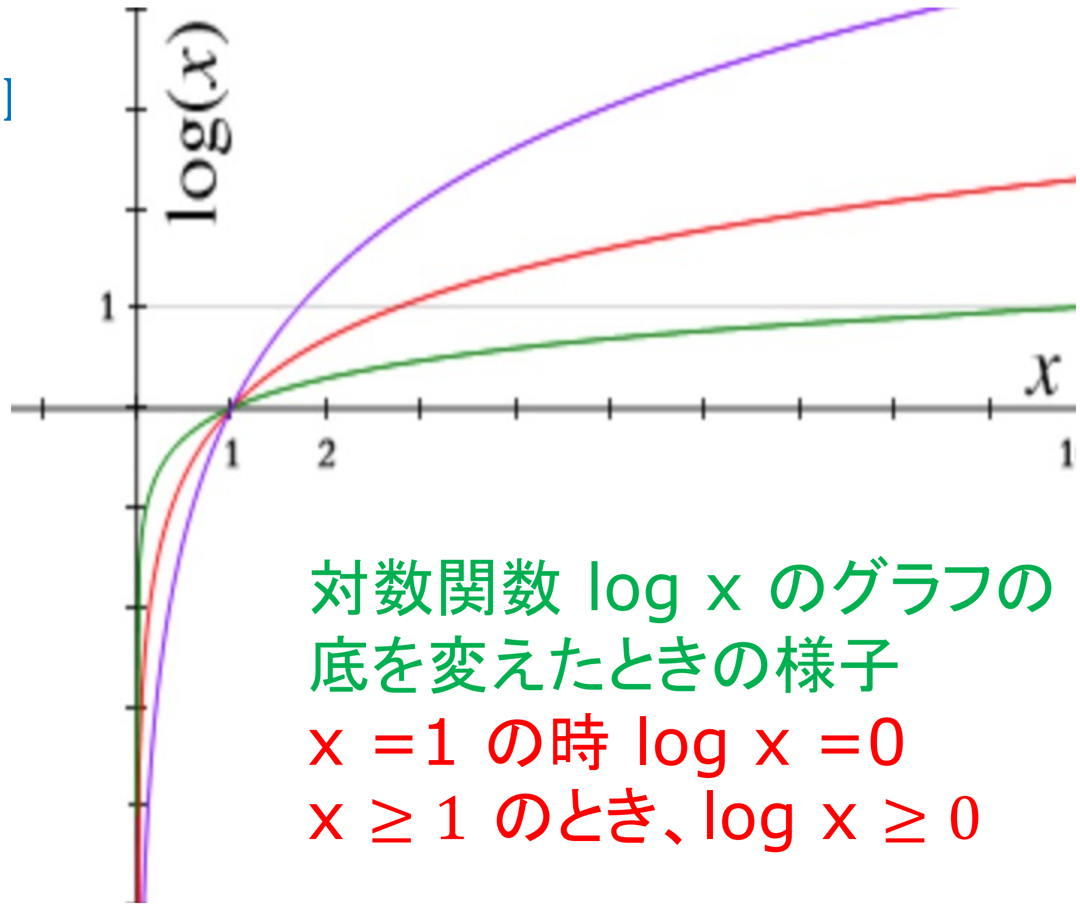
と変形できる。（ $-\log\ a = \log a^{-1} = \log \frac{1}{a}$ ）

$$\mathrm{H(X)} \geq 0$$

$$\mathrm{H(X)} = -\sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

p_iは確率だから、$0 \leq p_i \leq 1$ で $\frac{1}{p_i} \geq 1$ である。

対数の性質



対数関数 log x のグラフの
底を変えたときの様子

x =1 の時 log x =0

x ≥ 1 のとき、log x ≥ 0

$$\mathrm{H(X)} \geq 0$$

$$\mathrm{H(X)} = -\sum_i p_i \log_2 p_i = \sum_i p_i \log_2 \frac{1}{p_i}$$

p_iは確率だから、$0 \leq p_i \leq 1$ で $\frac{1}{p_i} \geq 1$ である。

この時、対数の性質から、$\log_2 \frac{1}{p_i} \geq 0$

よって、p_i も $\log_2 \frac{1}{p_i}$ も ≥ 0 であり $p_i \log_2 \frac{1}{p_i} \geq 0$

これから $\mathrm{H(X)} \geq 0$がわかる。

シャノン・エントロピーの性質について 2

$H(X) = 0$となる場合

H(X) = 0 となる場合

$$\mathrm{H(X)} = \sum_i p_i \log_2 \frac{1}{p_i}$$

が 0となるのは、$p_i = 0$ あるいは $\log_2 \frac{1}{p_i} = 0$ すなわち、$p_i = 1$ の場合である。

ただし、

$$\sum_i p_i = 1$$

である。

H(X) = 0 となる場合

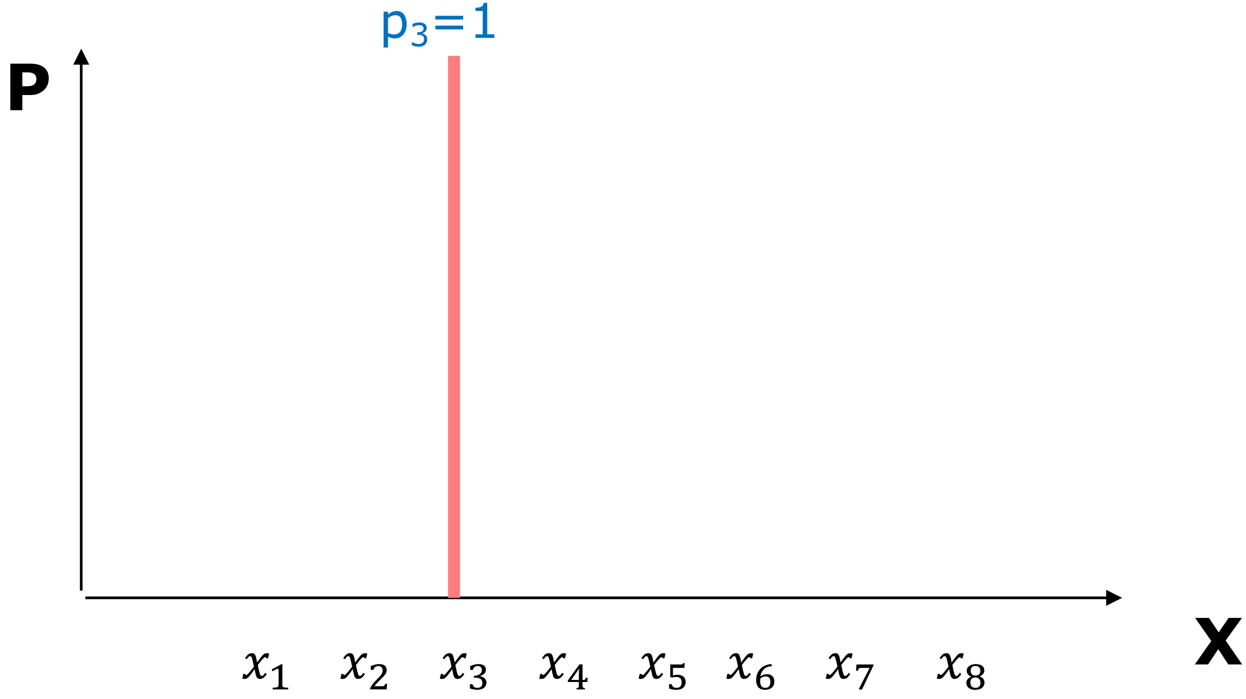
$$\mathrm{H(X)} = \sum_{i} p_i \log_2 \frac{1}{p_i}$$

が 0となるのは、$p_i = 0$ あるいは $\log_2 \frac{1}{p_i} = 0$ すなわち、$p_i = 1$ の場合である。

H(X) = 0 となるのは、ある一つの i についてのみ、$p_i = 1$ となり、それ以外の全ての jについて、$p_j = 0$ になる場合 である。

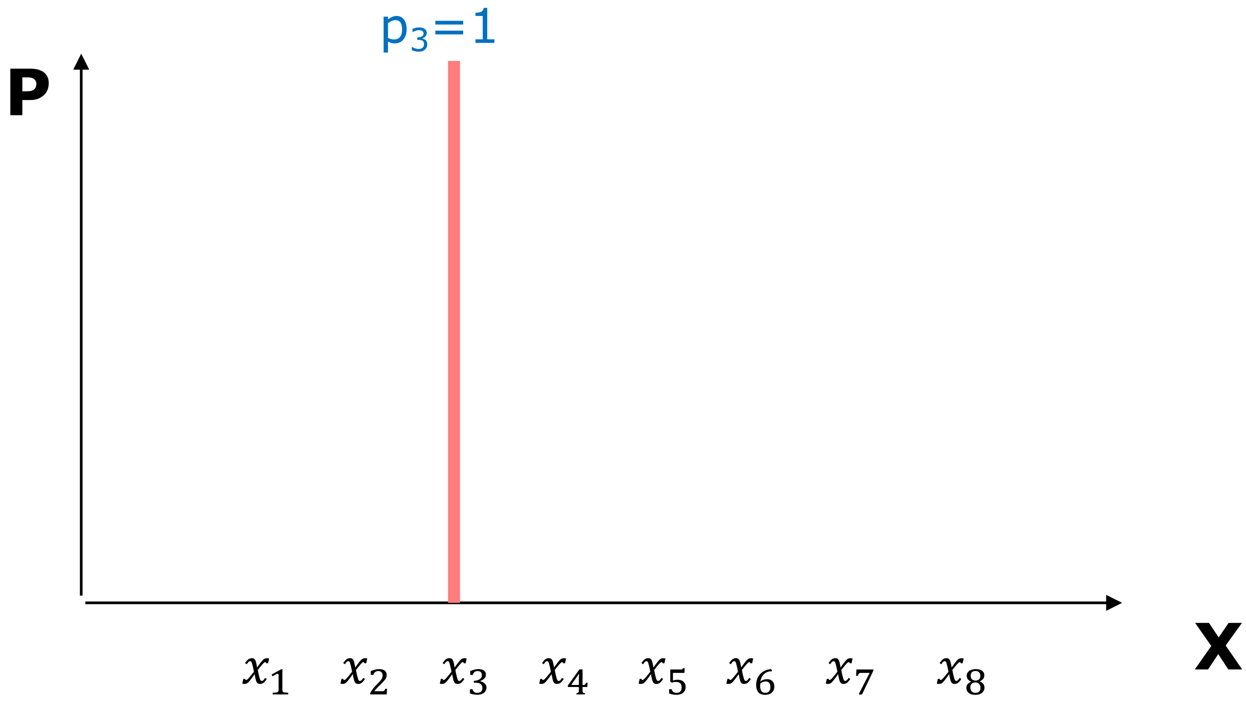
ある事象が確実に起こるとき、そのエントロピーはゼロになる。

一つのp_iのみが1の場合



この例では、Xがx_3の値を取る場合のみ、$p_3=1$である。
この時、H(X)=1･log(1/1)=1･log1=1･0=0
（ただし、0･log(1/0）=0とした）

一つのp_iのみが1の場合
それは、必ずx_iが起きることを意味する



ある事象xiが起きることの確率が1なら、それは必ずxiが起きることを意味する。その時、エントロピーの値はゼロになる。

確実な情報のエントロピーは、ゼロである。

シャノン・エントロピーの性質について 3

$H(X)$が最大となる場合

H(X)が最大の値をとる場合

ある事象が起きることが不確実なものになるにつれて、エントロピーは増大する。

n個の事象 X=$\{x_1, x_2,\ x_3, \ldots,\ \ x_n\}$ に対してX上の
確率分布　P=$\{p_1, p_2,\ p_3, \ldots,\ \ p_n\}$ が与えられているとしよう。
この時、H(X) ＝ H($p_1, p_2,\ p_3, \ldots,\ \ p_n$)と表すことにする。

H(X)が最大になるのは、どのx_i が選択されるか確実には分からない場合、すなわち、$p_1 = p_2 = p_3 = \cdots = p_n$の場合である。

H(X)が最大の値をとる場合

ある事象が起きることが不確実なものになるにつれて、エントロピーは増大する。

n個の事象 X={$x_1, x_2,\ x_3, \ldots,\ \ x_n$} に対してX上の
確率分布　P={$p_1, p_2,\ p_3, \ldots,\ \ p_n$} が与えられているとしよう。
この時、H(X) ＝ H($p_1, p_2,\ p_3, \ldots,\ \ p_n$)と表すことにする。

H(X)が最大になるのは、どのx_i が選択されるか確実には分からない場合、すなわち、$p_1 = p_2 = p_3 = \cdots = p_n$の場合である。
これは、$p_1 +\ p_2 + p_3 + \cdots + p_n = 1$ だから、全てのiについて
p_i=1/n の場合である。

H(X)が最大の値をとる場合

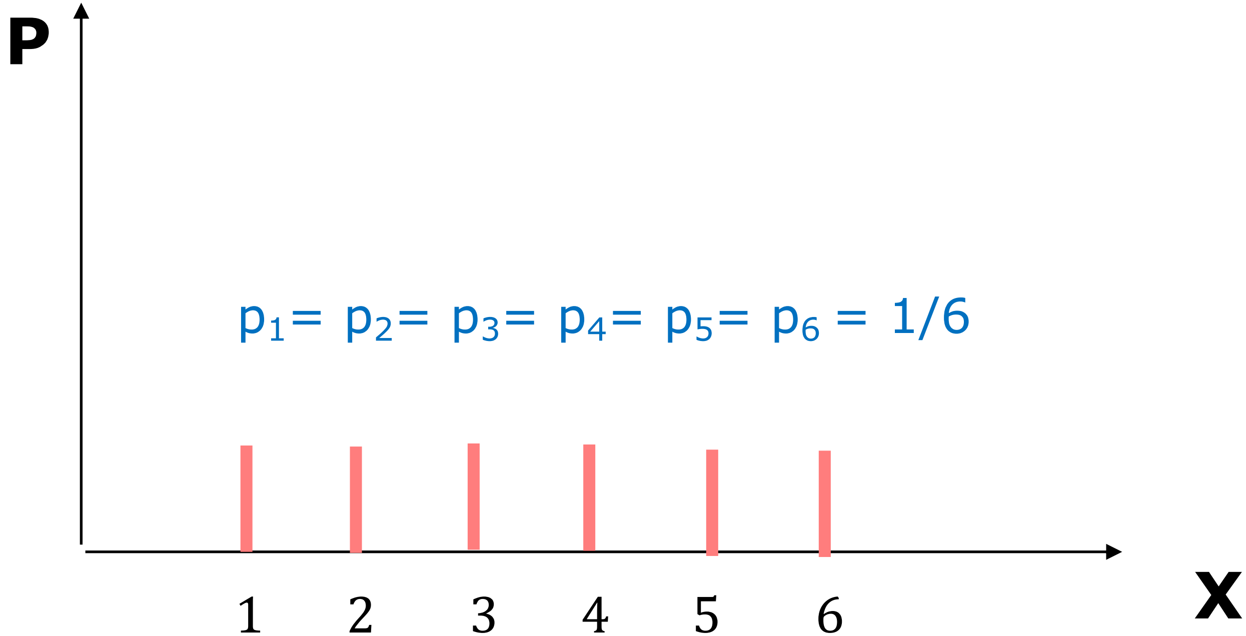
ある事象が起きることが不確実なものになるにつれて、エントロピーは増大する。

n個の事象 X=$\{x_1, x_2,\ x_3, \ldots,\ \ x_n\}$ に対してX上の
確率分布　P=$\{p_1, p_2,\ p_3, \ldots,\ \ p_n\}$ が与えられているとしよう。
この時、H(X) = H($p_1, p_2,\ p_3, \ldots,\ \ p_n$)と表すことにする。

H(X)が最大になるのは、どのx_i が選択されるか確実には分からない場合、すなわち、$p_1 = p_2 = p_3 = \cdots = p_n$の場合である。
これは、$p_1 + p_2 + p_3 + \cdots + p_n = 1$ だから、全てのiについて
p_i=1/n の場合である。この時、エントロピーの最大値は、

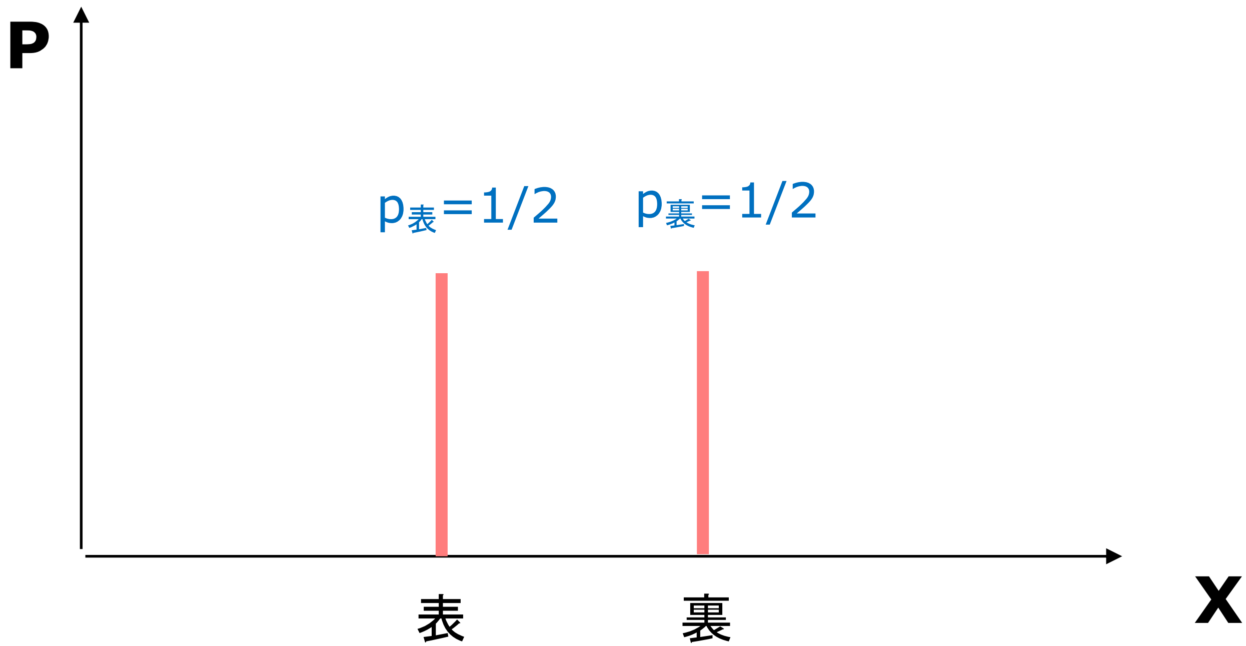
$$\mathrm{H(X)} = \sum_i p_i \log_2 \frac{1}{p_i} = \sum_i \frac{1}{n} \log_2 n = \log_2 n$$

サイコロの出る目の確率分布



p1=p2=p3=p4=p5=p6= 1/6 だから
H(X)=1/6･(log 6+log 6+log 6+log6+log 6+log 6)=log 6
= 1+log 3 bit

コイン・トスの確率分布

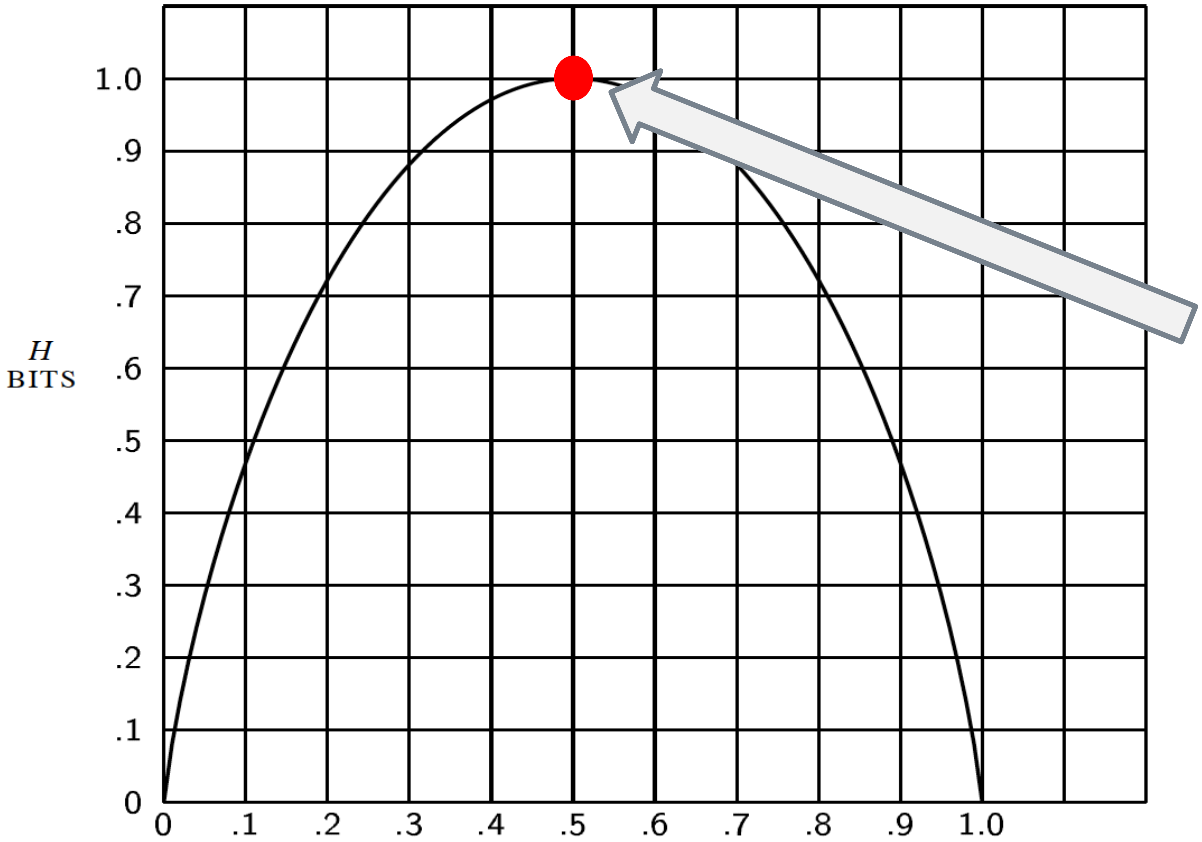


コインを投げた時、表が出る確率は1/2、裏が出る確率は1/2。

$H(X) = 1/2 \cdot \log 2 + 1/2 \cdot \log 2 = \log 2 = 1$

コイン・トスの確率分布のエントロピーは、1 bitである。

$p+q=1$ を満たす二項の確率分布とそのエントロピーのグラフ



$p=q=\frac{1}{2}$
の時に
エントロピーは
最大になる

$p+q=1$ の時の

$H(\mathrm{p},\mathrm{q})=-\mathrm{p}\log\mathrm{p}\;-\mathrm{q}\log\mathrm{q}=\mathrm{p}\log\left(\frac{1}{\mathrm{p}}\right)+\mathrm{q}\log(\frac{1}{\mathrm{q}})$ のグラフ

シャノン･エントロピーの性質について 4

エントロピーの単位としての bit

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、

H(1/2, 1/2) $= \log_2 2 = 1$

これをエントロピーの単位として bit と呼ぶ。

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、
$H(1/2,\ 1/2) = \log_2 2 = 1$
これをエントロピーの単位として bit と呼ぶ。
bit は、二つの状態を持つものの最大エントロピーである。

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、
H(1/2, 1/2) = $\log_2 2 = 1$
これをエントロピーの単位として bit と呼ぶ。
bit は、二つの状態を持つものの最大エントロピーである。

X={0,1,2,3}の時、H(X)の最大エントロピーは
H(1/4, 1/4, 1,4, 1/4) = $\log_2 4 = 2\log_2 2 = 2\,bit$
四つの状態を持つものの最大エントロピーは 2bit である。

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、

H(1/2, 1/2) $= \log_2 2 = 1$

これをエントロピーの単位として bit と呼ぶ。

bit は、二つの状態を持つものの最大エントロピーである。

X={0,1,2,3}の時、H(X)の最大エントロピーは

H(1/4, 1/4, 1,4, 1/4) $= \log_2 4 = 2\log_2 2 = 2\,bit$

四つの状態を持つものの最大エントロピーは 2bit である。

一般に 2^nの状態を持つものの最大エントロピーは

$$\log_2 2^n = n\log_2 2 = n\,bit$$

シャノンエントロピーの単位としての bit

X={0,1} の時、H(X)の最大エントロピーは、
H(1/2, 1/2) $= \log_2 2 = 1$
これをエントロピーの単位として bit と呼ぶ。
bit は、二つの状態を持つものの最大エントロピーである。

X={0,1,2,3}の時、H(X)の最大エントロピーは
H(1/4, 1/4, 1,4, 1/4) $= \log_2 4 = 2\log_2 2 = 2\ bit$
四つの状態を持つものの最大エントロピーは 2bit である。

一般に 2^nの状態を持つものの最大エントロピーは

$$\log_2 2^n = n\log_2 2 = n\ bit$$

これは、我々のbit についての直観と一致している。

第二部

シャノンが考えたことを振り返る

第二部 Agenda
シャノンが考えたことを振り返る

- なぜ情報の尺度に対数が現れるのか？
- 選択と不確かさとエントロピー
  -- エントロピーが満たすべき条件を考える
- エントロピーが満たすべき条件とそれから導かれる関数等式
- シャノンは、どのようにエントロピーの式を導出したのか (1)
- シャノンは、どのようにエントロピーの式を導出したのか (2)

シャノンが考えたこと１

なぜ情報の尺度に対数が現れるのか？

以前に、ボルツマンのエントロピーからシャノンのエントロピーが導かれることを見てきたことがある。

2021/05/29 マルレク基礎「情報とエントロピー入門」の
https://www.marulabo.net/docs/info-entropy/
「ボルツマンのエントロピーからシャノンのエントロピーを導く」
https://drive.google.com/file/d/1bXyIxN7ASmfkBM8QRle69U56C_fo-Vak/view?usp=sharing を参照されたい。

しかし、シャノンはこのようにして情報量としてのエントロピーを導いたわけではない。

シャノンは、通信システムの分析を通じて、情報量＝エントロピーの概念に到達した。ここでは、彼が、どういうことを考えていたかをあらためて、振り返ってみよう。

情報の尺度を定義する
「通信の数学理論」

A Mathematical Theory of Communication
C. E. SHANNON
http://www.qsl.net/n9zia/pdf/shannon1948.pdf
1948年

一般的な通信システムの図式

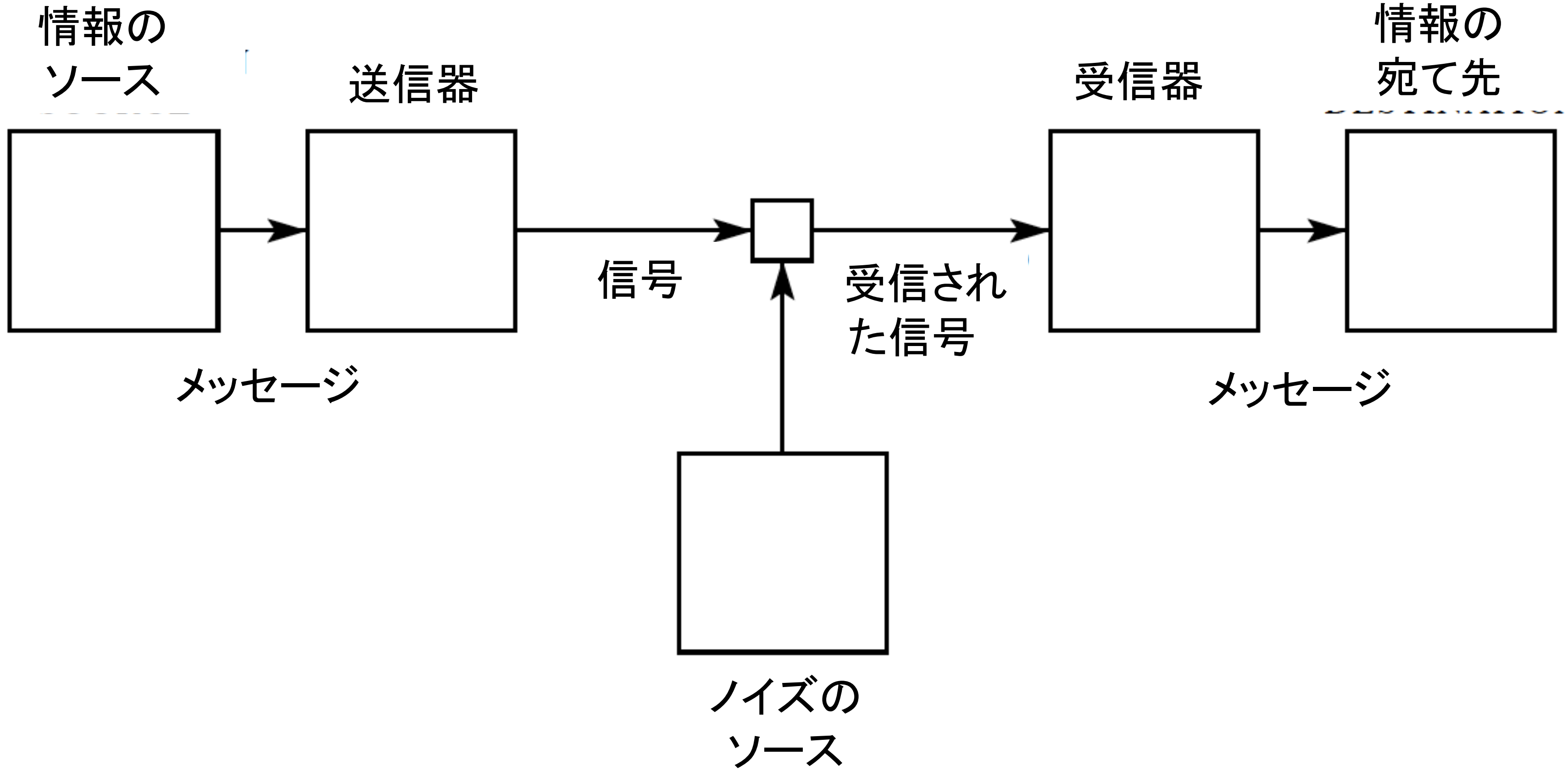
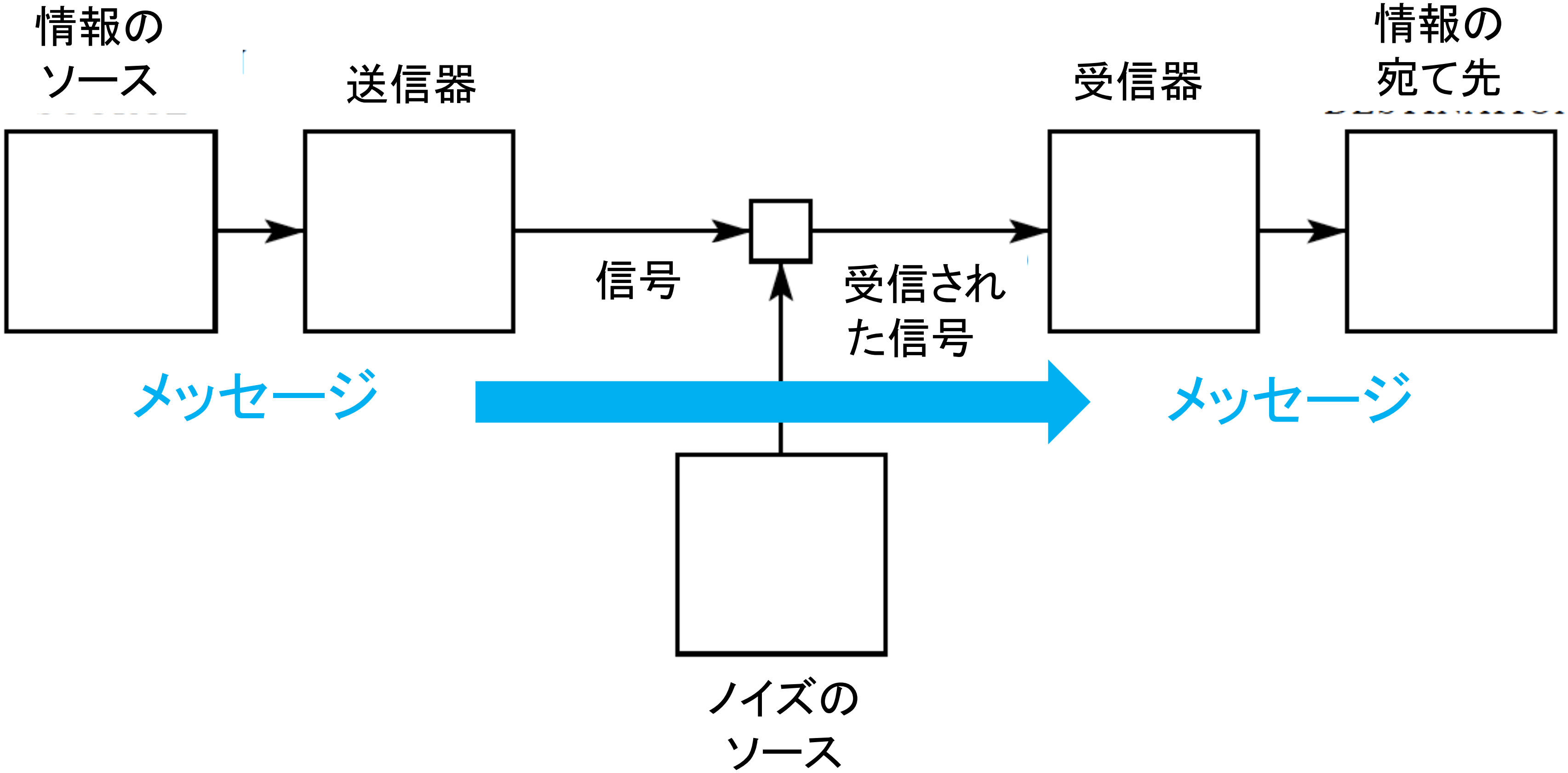


Fig. 1—Schematic diagram of a general communication system.

一般的な通信システムの図式と
通信の基本問題



通信の基本的な問題は、情報のソースで選択されたメッセージを、正確であれ近似的であれ、情報の宛て先で再生産することである。

メッセージの「意味」は、重要ではない

多くの場合、そのメッセージは意味を持っている。すなわち、そのメッセージは、あるシステムに従って確かな物理的実体あるいは概念的実体を、参照するかそれに関連している。ただ、こうした通信の意味論的な諸側面は、工学的問題とは無関係である。

メッセージの「意味」は、重要ではない

多くの場合、そのメッセージは意味を持っている。すなわち、そのメッセージは、あるシステムに従って確かな物理的実体あるいは概念的実体を、参照するかそれに関連している。ただ、こうした通信の意味論的な諸側面は、工学的問題とは無関係である。

重要な側面は、実際のメッセージは可能なメッセージの集合の中から選択された一つのメッセージだと言うことである。

メッセージの「意味」は、重要ではない

多くの場合、そのメッセージは意味を持っている。すなわち、そのメッセージは、あるシステムに従って確かな物理的実体あるいは概念的実体を、参照するかそれに関連している。ただ、こうした通信の意味論的な諸側面は、工学的問題とは無関係である。

重要な側面は、実際のメッセージは可能なメッセージの集合の中から選択された一つのメッセージだと言うことである。

通信システムは、実際に選択されたメッセージだけに対してではなく、全ての可能な選択に対して機能するように設計されていなければならない。なぜなら、設計の時点では、どのメッセージが選ばれるかは、わからないからである。

情報の尺度として「対数」を利用する

もしも、選択されるメッセージの集合の要素の数が有限であるなら、この集合から一つのメッセージが選ばれた時、その選択はすべて同じようにありうるので、この数あるいはこの数の単調な関数はすべて、生成されたこの情報の尺度と見なすことができる。

情報の尺度として「対数」を利用する

もしも、選択されるメッセージの集合の要素の数が有限であるなら、この集合から一つのメッセージが選ばれた時、その選択はすべて同じようにありうるので、この数あるいはこの数の単調な関数はすべて、生成されたこの情報の尺度と見なすことができる。

Hartlyが指摘したように、この関数として最も自然なのは、対数を選ぶことである。

この定義は、メッセージの統計的性質の影響や、メッセージの値が連続的なものである場合など、よく考えた上で一般化されねばならないのだが、我々は全ての場合で、情報の尺度として、本質的には対数的な尺度を利用するだろう。

対数的尺度が便利であるいくつかの理由

工学的に重要な、時間。帯域、リレーの数といったパラメーターは、可能な数の対数について、線形に変化する傾向がある。

例えば、リレーを一個追加すると、可能な状態の数は2倍になる。元の状態を数をNとすれば、リレー一個を追加した時可能な状態の数は2Nとなる。状態の数 N→2Nの変化は、2を底とする対数で考えれば、$\log N \rightarrow \log 2N = \log 2 + \log N = 1 + \log N$ である。

対数的尺度が便利であるいくつかの理由

工学的に重要な、時間。帯域、リレーの数といったパラメーターは、可能な数の対数について、線形に変化する傾向がある。

例えば、リレーを一個追加すると、可能な状態の数は2倍になる。元の状態を数をNとすれば、リレー一個を追加した時可能な状態の数は2Nとなる。状態の数 N→2Nの変化は、2を底とする対数で考えれば、log N → log 2N = log 2 + log N = 1 + log N である。

時間が2倍になれば、可能なメッセージの数は、元の可能なメッセージの数の2乗になるが、対数で考えれば、2倍になる。

対数的尺度が便利であるいくつかの理由

我々は、直感的には、ある実体の尺度を、共通の基準と線形で比較する。

二枚のパンチカードは一枚のカードより、2倍の情報容量を持つと感じる。

二つの同じ通信チャンネルがあれば、一つの通信チャンネルより、2倍の情報容量を持つと感じる。

情報の尺度としての bit

対数の底の選択は、情報の尺度の単位の選択に対応している。
対数の底に 2を選んだ時、その単位は bit と呼ばれる。

二つの状態を取るデバイスは、log 2 ＝ 1 bitの情報を蓄える。
こうしたデバイスががN個あった時、N個のデバイスが取りうる状態の総数は、2^Nである。log 2^N = N だから、これらのデバイスは N bitの情報を蓄えることになる。

シャノンが考えたこと 2

選択と不確かさとエントロピー

選択と不確かさとエントロピー

我々は、離散的な情報源をマルコフ過程として表現してきた。

我々は、ある意味で、どれほどの情報がこうしたプロセスによって生成されたかを、さらには、どのような割合で情報が生成されたかを測定する量を定義できるだろうか？

選択と不確かさとエントロピー

我々は、離散的な情報源をマルコフ過程として表現してきた。

我々は、ある意味で、どれほどの情報がこうしたプロセスによって生成されたかを、さらには、どのような割合で情報が生成されたかを測定する量を定義できるだろうか？

それが生起する確率が、p1, p2, … , pnの可能なイベントの集合が与えられたとしよう。これらの確率は知られているが、我々が知っていることのすべては、どれかのイベントが起きるだろうと言うことのみである。

選択と不確かさとエントロピー

我々は、離散的な情報源をマルコフ過程として表現してきた。

我々は、ある意味で、どれほどの情報がこうしたプロセスによって生成されたかを、さらには、どのような割合で情報が生成されたかを測定する量を定義できるだろうか？

それが生起する確率が、p1, p2, … , pnの可能なイベントの集合が与えられたとしよう。これらの確率は知られているが、我々が知っていることのすべては、どれかのイベントが起きるだろうと言うことのみである。

そのイベントを選び出すのに、どれほどの「選択」が含まれているのか、あるいは、その出力について、我々はいかに不確かなのかを測定する尺度を、我々は見つけ出すことができるだろうか？

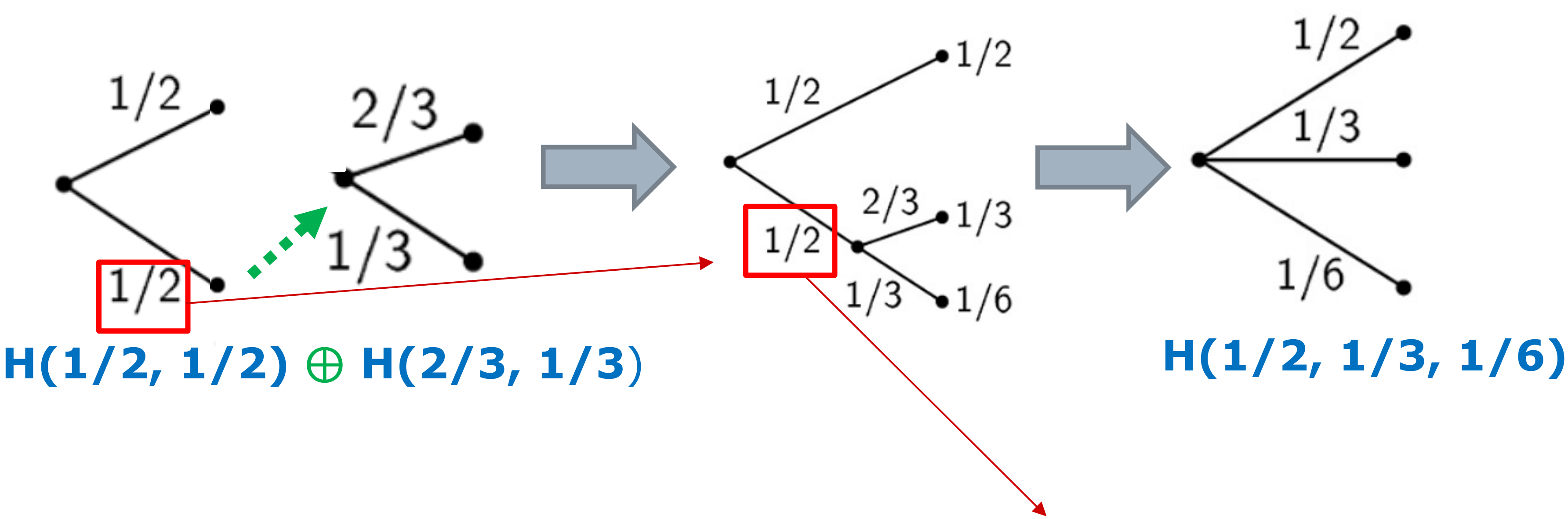
Hが満たすべき条件

もし、そうした尺度があるとして、それを H(p1, p2, … ,pn) と表そう。この時、この関数Hが、次のような性質を満たすことを要求するのは合理的である。

1. H は、pi について連続的である。
2. もし全てのpiが等しく、pi ＝ 1/n であるとすれば、Hは、nについて単調増加の関数である。
   （同じようように見えるイベントが与えられ、より多くの可能なイベントが存在する時、より多くの選択、より多くの不確かさが存在することになる。）
3. もし選択が、連続的に行われる二つの選択に分解されるのなら、求めるHは、個々の選択のHの値の、重みづけられた和になるべきである。

先の条件 3. の説明

H(1/2, 1/2)で表される選択と、H(2/3, 1/3)で表される選択を、次のように連続して行なったとする。



$$H(1/2,\ 1/3,\ 1/6) = H(1/2,\ 1/2) + 1/2\ H(2/3,\ 1/3)$$

シャノンが証明したこと

先の 1,2,3の条件を満たす関数 Hは、ただ一つで、次の形をしている。（ただし、Kは正の定数）

$$\mathrm{H}(p_1, p_2, \ldots, p_n) = -K \sum_i p_i \log p_i$$

シャノンが考えたこと 3

エントロピーが満たすべき条件と
それから導かれる関数等式

シャノンが考えた
エントロピーHが満たすべき三つの条件

シャノンは、エントロピーHが、次の条件を満たすべきだと考えた。

1. H は、pi について連続的である。
2. もし全てのpiが等しく、pi ＝ 1/n であるとすれば、Hは、nについて単調増加の関数である。
   （同じように見えるイベントが与えられ、より多くの可能なイベントが存在する時、より多くの選択、より多くの不確かさが存在することになる。）
3. もし選択が、連続的に行われる二つの選択に分解されるのなら、求めるHは、個々の選択のHの値の、重みづけられた和になるべきである。

条件 3. から導かれる式 $A(s^m) = mA(s)$

$H\left(\frac{1}{n}, \frac{1}{n}. \cdots, \frac{1}{n}\right) = A(n)$ とする。
この時、シャノンは、条件 3. から、次の式が成り立つという。

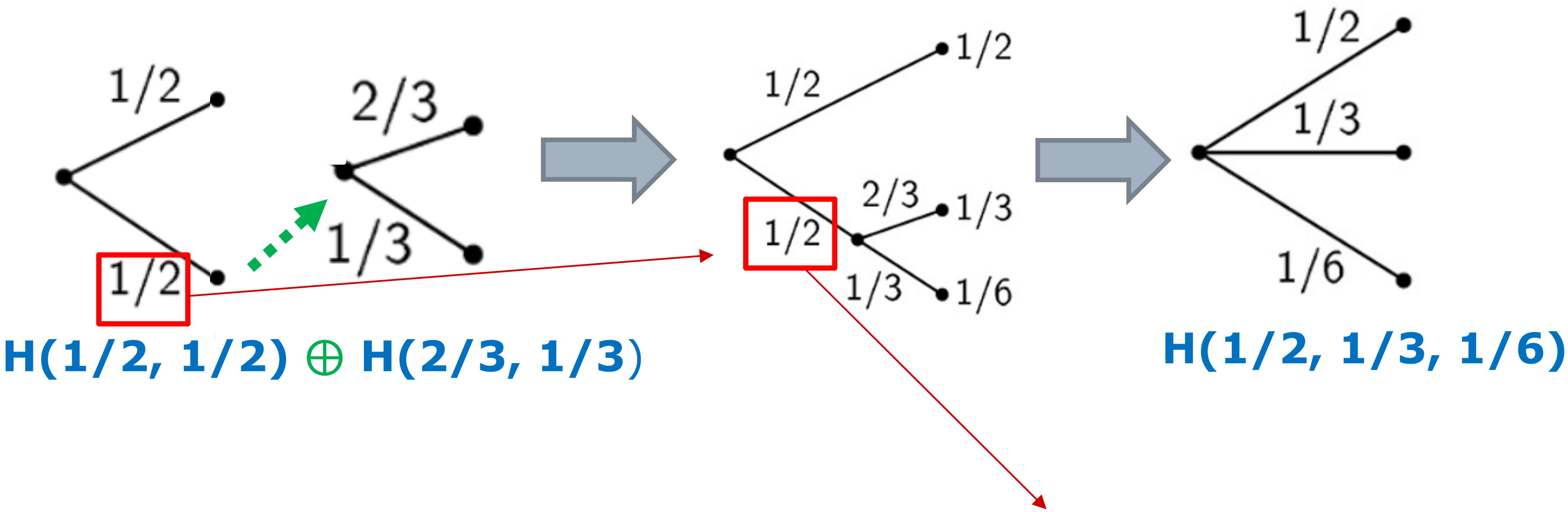
$$A(s^m) = mA(s)$$

この式は、次のセッションで見るシャノンのエントロピーの式の導出で、基本的な役割りを果たす。

以下、この式が成り立つことを、みていこう。

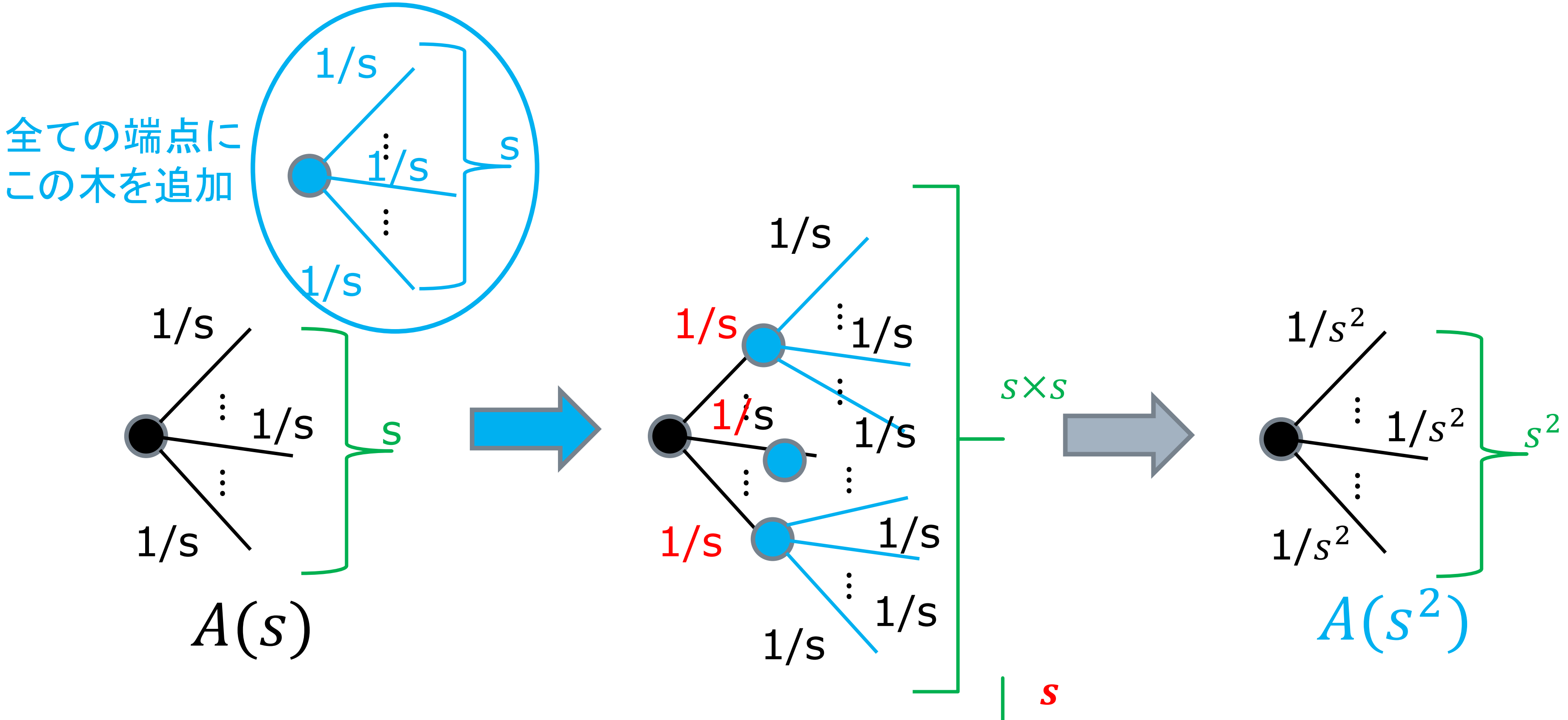
先の条件 3. で、シャノンのあげた例

H(1/2, 1/2)で表される選択と、H(2/3, 1/3)で表される選択を、次のように連続して行なったとする。



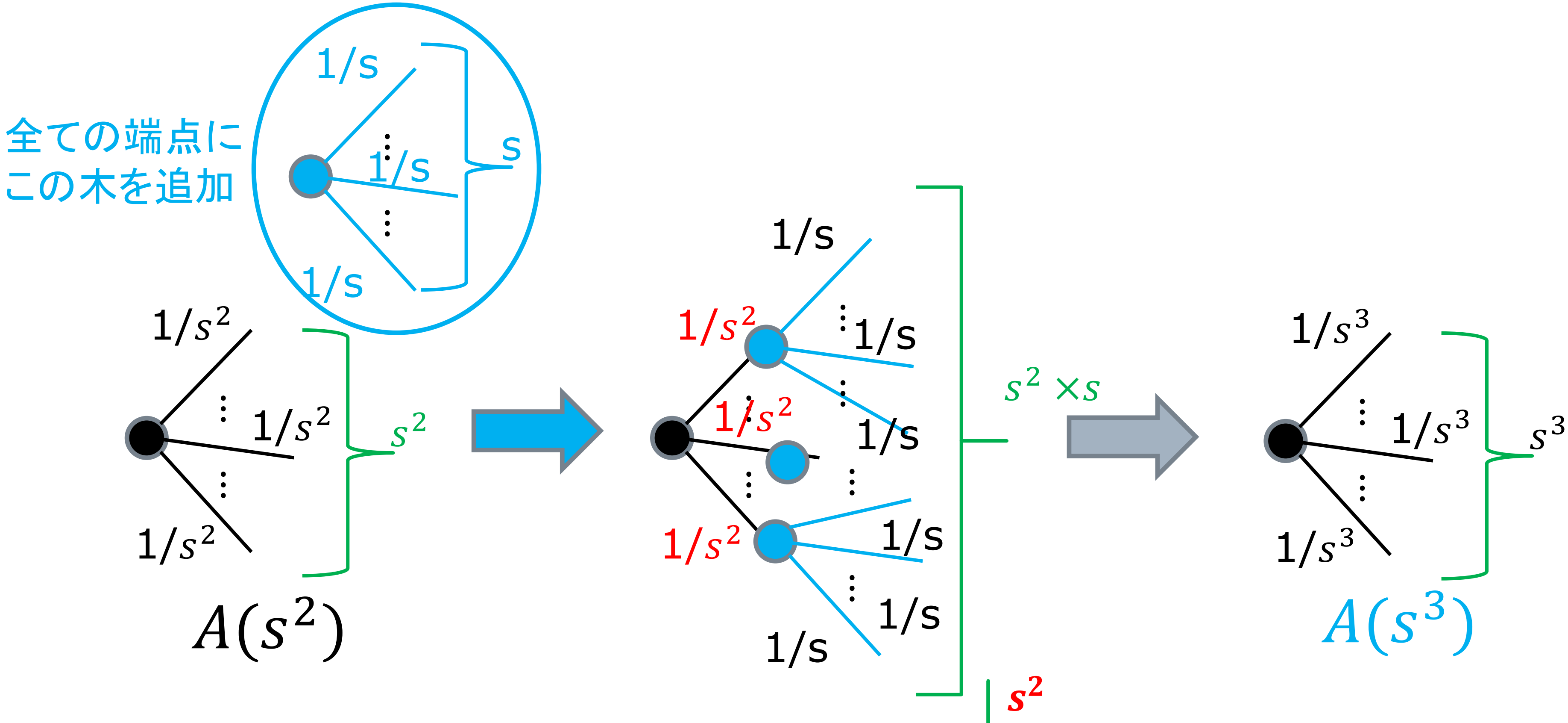
$$H(1/2, 1/3, 1/6) = H(1/2, 1/2) + 1/2\ H(2/3, 1/3)$$

$m=2$の時、$A(s^2)=2A(s)$



$$A(s^2) = A(s) + \frac{1}{s}A(s) + \frac{1}{s}A(s) + \cdots + \frac{1}{s}A(s)$$
$$= A(s) + A(s) = 2A(s)$$

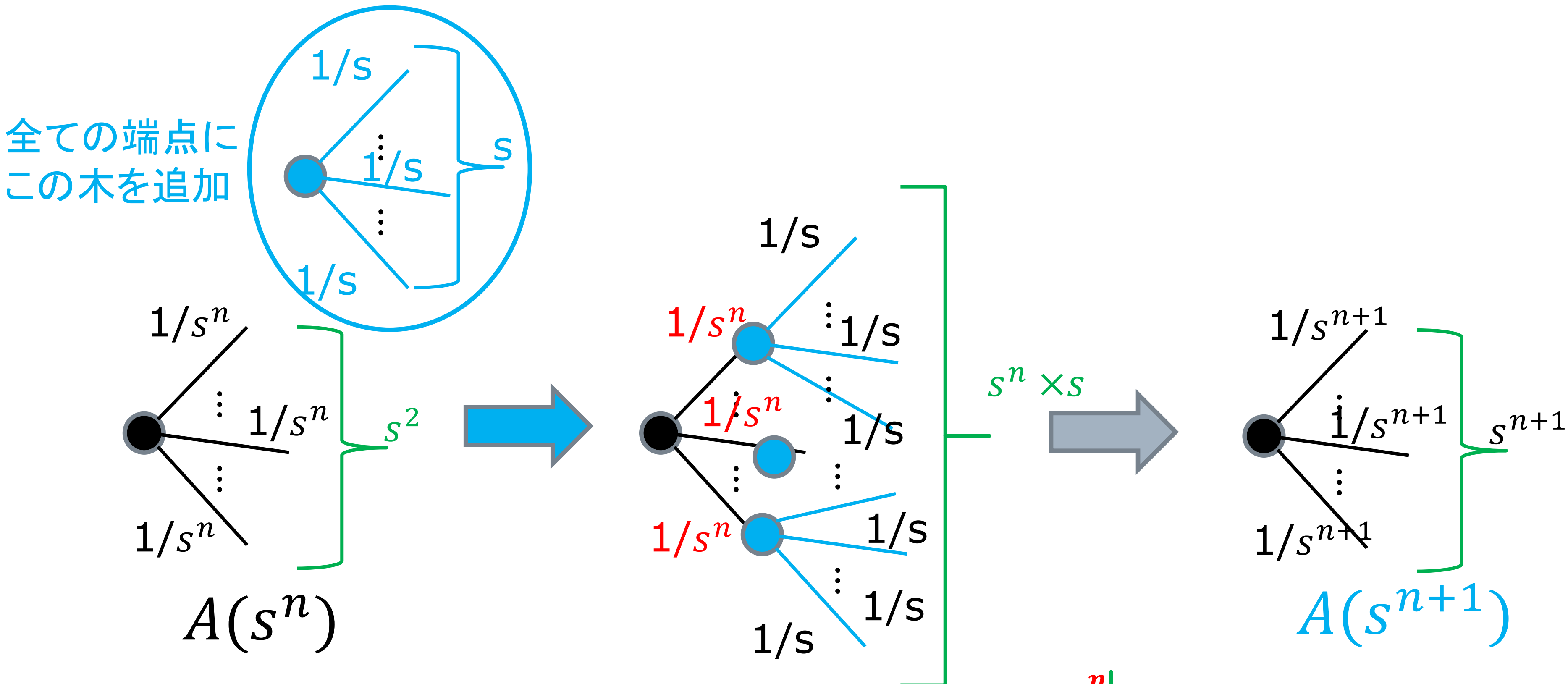
$m=3$の時、$A(s^3)=3A(s)$



$$A(s^3)=A(s^2)+\frac{1}{s^2}A(s)+\frac{1}{s^2}A(s)+\cdots+\frac{1}{s^2}A(s)$$

$$=2A(s)+A(s)=3A(s)$$

$m=n+1$の時、$A(s^{n+1})=(n+1)A(s)$



$$A(s^{n+1}) = A(s^n) + \overbrace{\frac{1}{s^n}A(s) + \frac{1}{s^n}A(s) + \cdots + \frac{1}{s^n}A(s)}^{s^n}$$
$$= nA(s) + A(s) = (n+1)A(s)$$

シャノンが考えたこと 4

シャノンは、どのように
エントロピーの式を導出したのか (1)

シャノンの証明の概略

$H\left(\frac{1}{n}, \frac{1}{n}. \cdots, \frac{1}{n}\right) = A(n)$ とする。条件 (3) から、次の式を導く。

$A(s^m) = mA(s)$ (この式については、前回証明した)

- 証明の前半の課題(今回のセッション)

$A(s^m) = mA(s)$ を利用して、$A(t) = K \log t$ を導く

- 証明の後半(次回のセッション)

再び、条件 (3) と 先の$A(t) = K \log t$ から次の式を証明する

$$K \log \sum n_i = H(\mathrm{p}_1 \cdots p_n) + K \sum p_i \log n_i$$

この式から、エントロピーの式を導く。

$$\mathrm{H} = -K \sum_i p_i \log p_i$$

$A(s^m) = mA(s)$ から、$A(t) = K \log t$を導く

$H\left(\frac{1}{n}, \frac{1}{n}. \cdots, \frac{1}{n}\right) = A(n)$ とする。条件 (3) から、同じ確率をもつs^mからの選択を、同じ確率を持つ s からの連続したm個の選択に分解できる。次の式を得る。

$$A(s^m) = mA(s)$$

同様に、

$$A(t^n) = nA(t)$$

任意の大きさのnを選べるので、次の式を満たす mを見つけることができる。

$$s^m \leq t^n < s^{\{m+1\}}$$

$s^m \leq t^n < s^{\{m+1\}}$ なる $n.m$ が存在する

両辺のlogをとって、

$$m \log s \leq n \log t \leq (m+1) \log s$$

両辺を$n \log s$ で割ると、

$$\frac{m}{n} \leq \frac{\log t}{\log s} \leq \frac{m}{n} + \frac{1}{n}$$

すなわち、

$$\left| \frac{m}{n} - \frac{\log t}{\log s} \right| < \varepsilon$$

ε は、任意の小さな数である。

単調性と、$A(s^m) = mA(s)$

ここで、単調性の性質と先に見た$A(s^m) = mA(s)$を使う。

$$A(s^m) \le A(t^n) \le A(s^{m+1})$$
$$mA(s) \le nA(t) \le (m+1)A(s)$$

両辺を $nA(s)$ で割ると、

$$\frac{m}{n} \le \frac{A(t)}{A(s)} \le \frac{m}{n} + \frac{1}{n}$$

すなわち

$$\left|\frac{m}{n} - \frac{A(t)}{A(s)}\right| < \varepsilon$$

$$\left|\frac{m}{n} - \frac{A(t)}{A(s)}\right| < \varepsilon$$

また、先に見たように

$$\left|\frac{m}{n} - \frac{\log t}{\log s}\right| < \varepsilon$$

これから、

$$\left|\frac{A(t)}{A(s)} - \frac{\log t}{\log s}\right| < 2\varepsilon$$

こうして、

$$A(t) = K \log t$$

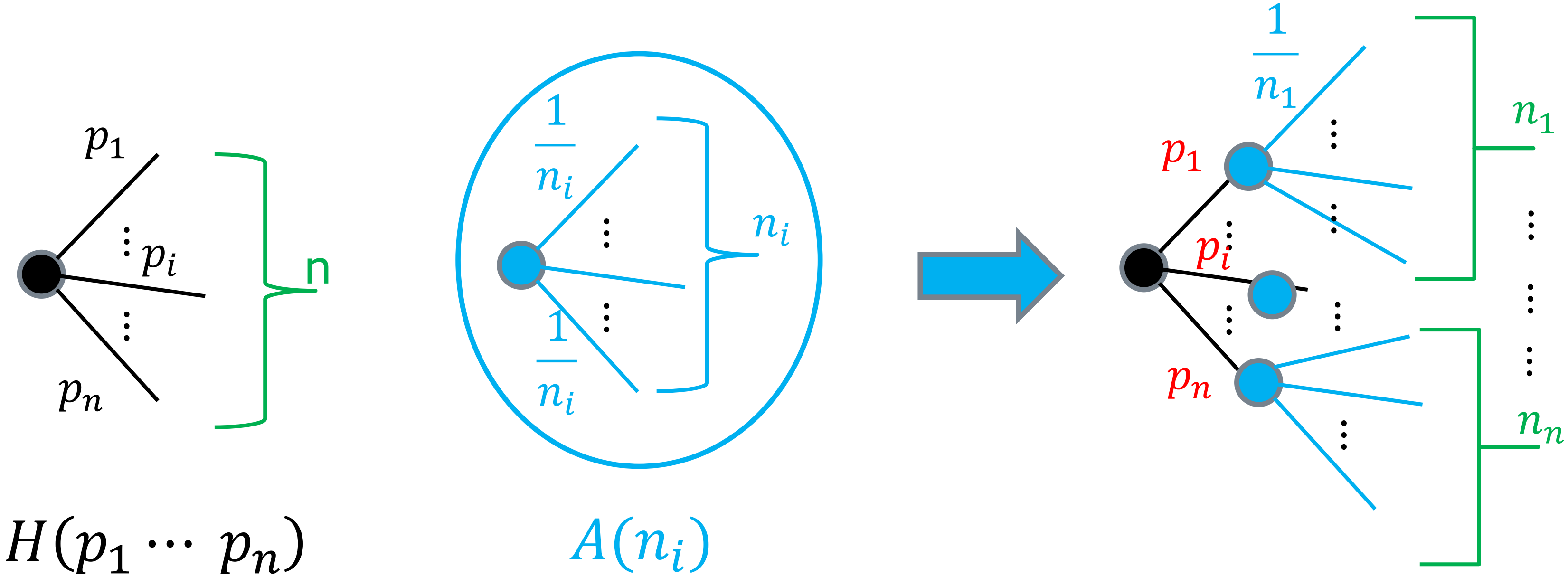
ここに、Kは (2).を満たすので正。

シャノンが考えたこと 5

シャノンは、どのように
エントロピーの式を導出したのか (2)

$$K\log\sum n_i = H(\mathrm{p}_1\cdots p_n) + K\sum p_i \log n_i$$

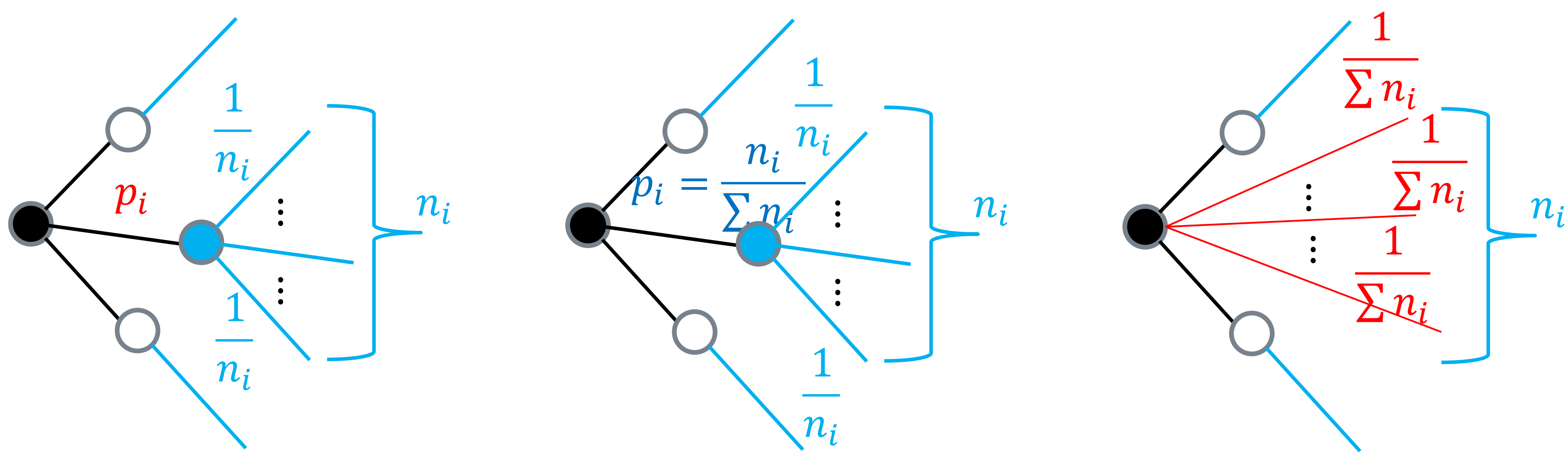
次のような木 $H(\mathrm{p}_1\cdots p_n)$の、各i 番目の端点に、
次の木 $A(n_i)$ を接続する。



新しい木の枝の数は、$n_1+n_2+\cdots+n_n=\sum n_i$ である。

$$p_i = \frac{n_i}{\sum n_i}$$

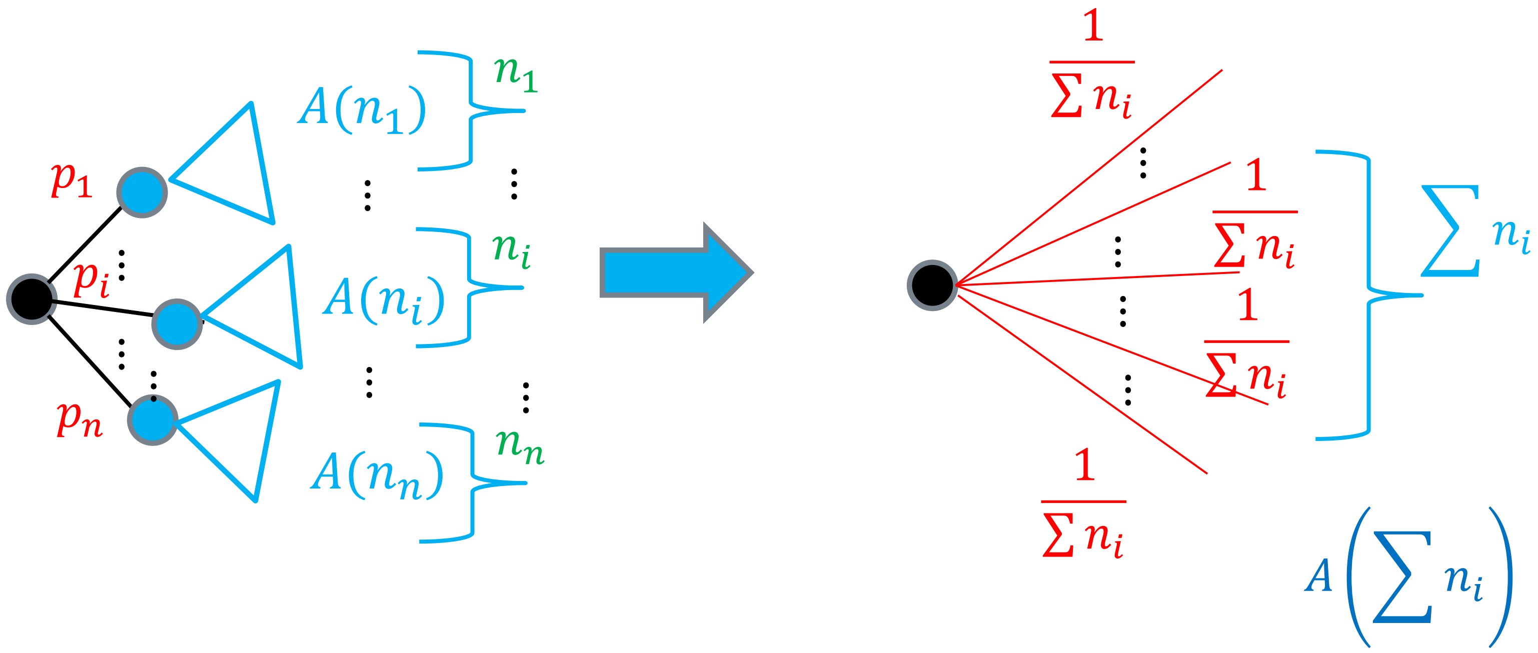
元の木 $H(\mathrm{p}_1 \cdots p_n)$の、i 番目の端点で、
何が起きているかを、少し詳しく見てみよう。



$p_i = \frac{n_i}{\sum n_i}$ を使うと、元の木の根から
新しい木の端点までの確率は、次の
式で求められる。

$$p_i \times \frac{1}{n_i} = \frac{n_i}{\sum n_i} \times \frac{1}{n_i} = \frac{1}{\sum n_i}$$

$A\left(\sum n_i\right)$の木ができる！



$$A\left(\sum n_i\right) = H(\mathrm{p}_1 \cdots\ p_n) + \mathrm{p}_1\mathrm{A}(\mathrm{n}_1) + \mathrm{p}_2\mathrm{A}(\mathrm{n}_2) + \cdots + p_n A(n_n)$$

$$A(t) = Klog\ t$$

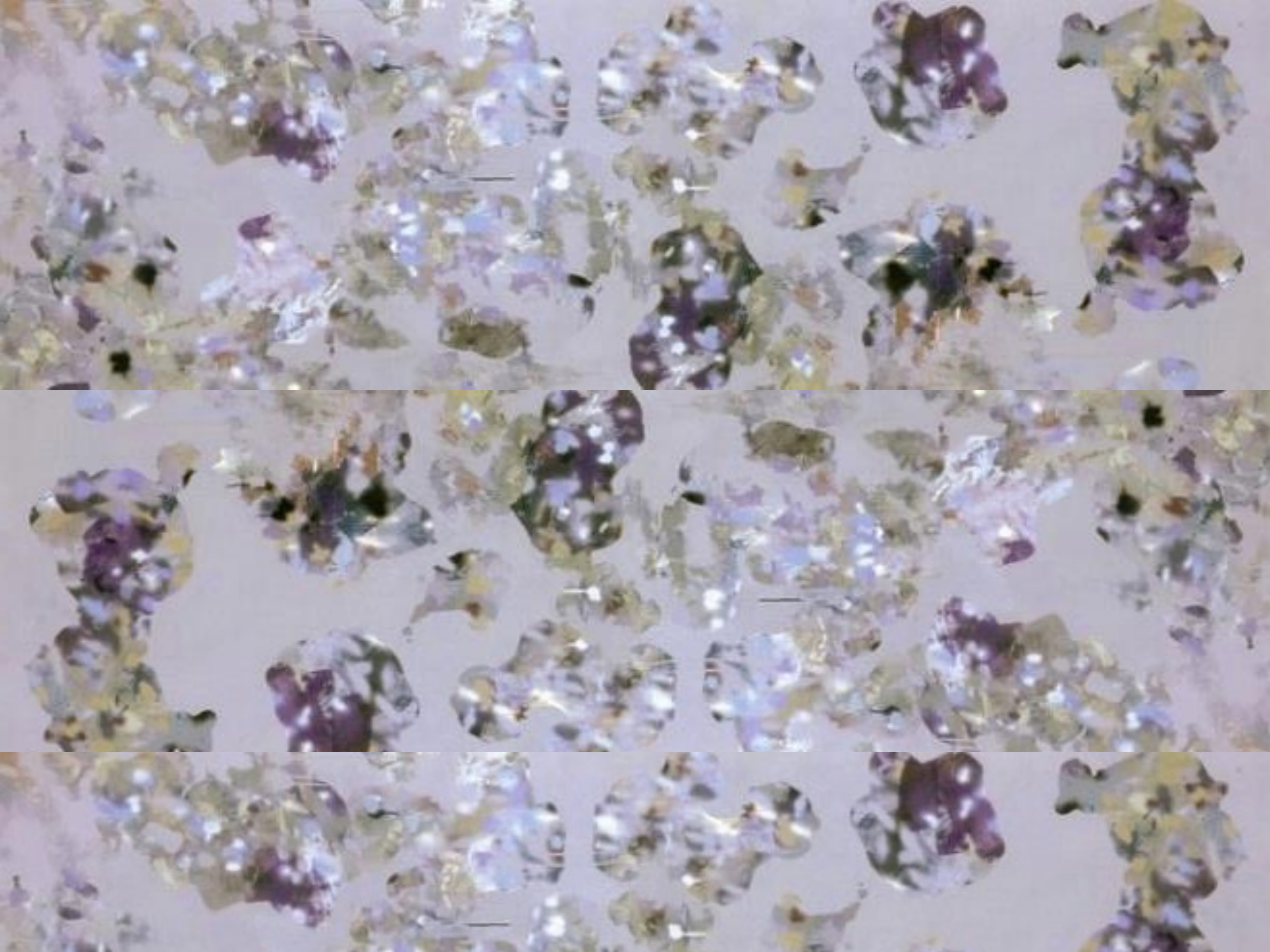
$$A\left(\sum n_i\right) = H(\mathrm{p}_1 \cdots p_n) + \mathrm{p}_1\mathrm{A}(\mathrm{n}_1) + \mathrm{p}_2\mathrm{A}(\mathrm{n}_2) + \cdots + p_n A(n_n)$$

先に得た $A(t) = Klog\ t$ を使うと、この式は、次のように変型できる。

$$K\log\sum n_i = H(\mathrm{p}_1 \cdots p_n) + K\sum p_i \log n_i$$

$$H = K\left[\sum p_i \log \sum n_i - \sum p_i \log n_i\right] \qquad \sum p_i = 1$$

$$= -K\sum p_i \log \frac{n_i}{\sum n_i} = -K\sum p_i \log p_i$$



第三部

Faddeev-LeinsterのChain ルール

第三部 **Agenda**
Faddeev-LeinsterのChain ルール

- Chainルールとは何か？
- Chainルールで
  シャノン・エントロピーを特徴づける
- シャノン・エントロピーは
  Chainルールを満たすことを確かめる

Chainルールとは何か？

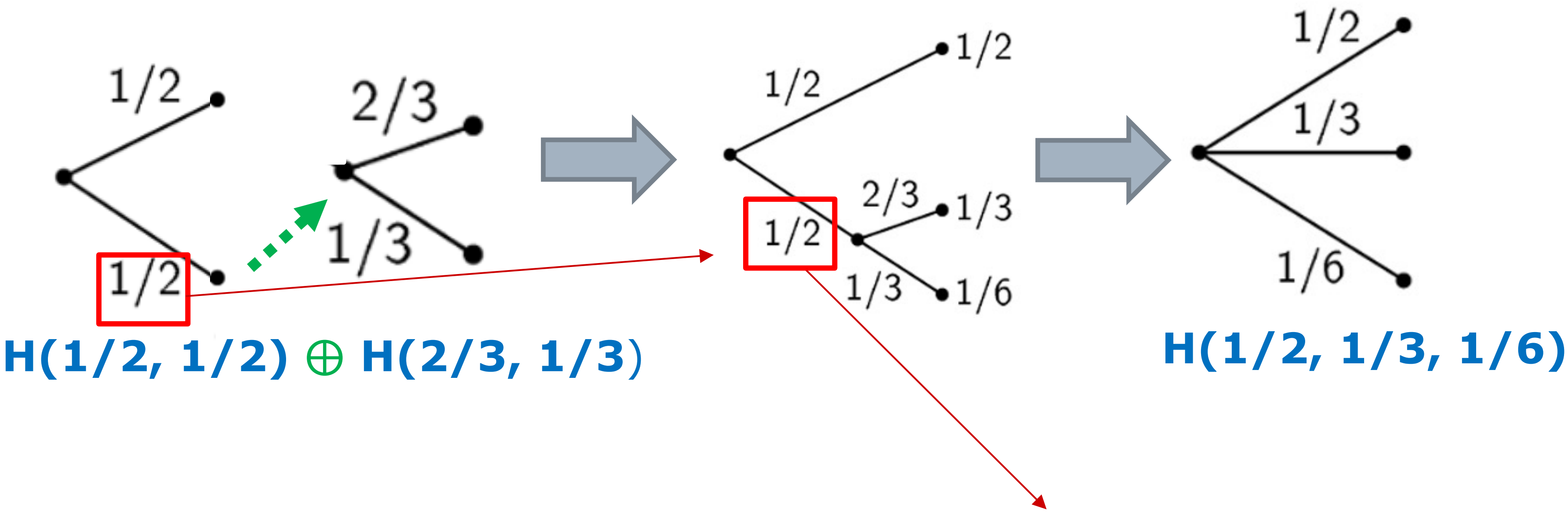
シャノンが考えた
エントロピーHが満たすべき三つの条件

シャノンは、エントロピーHが、次の条件を満たすべきだと考えた。

1. H は、pi について連続的である。
2. もし全てのpiが等しく、pi ＝ 1/n であるとすれば、Hは、nについて単調増加の関数である。
（同じように見えるイベントが与えられ、より多くの可能なイベントが存在する時、より多くの選択、より多くの不確かさが存在することになる。）
3. もし選択が、連続的に行われる二つの選択に分解されるのなら、求めるHは、個々の選択のHの値の、重みづけられた和になるべきである。

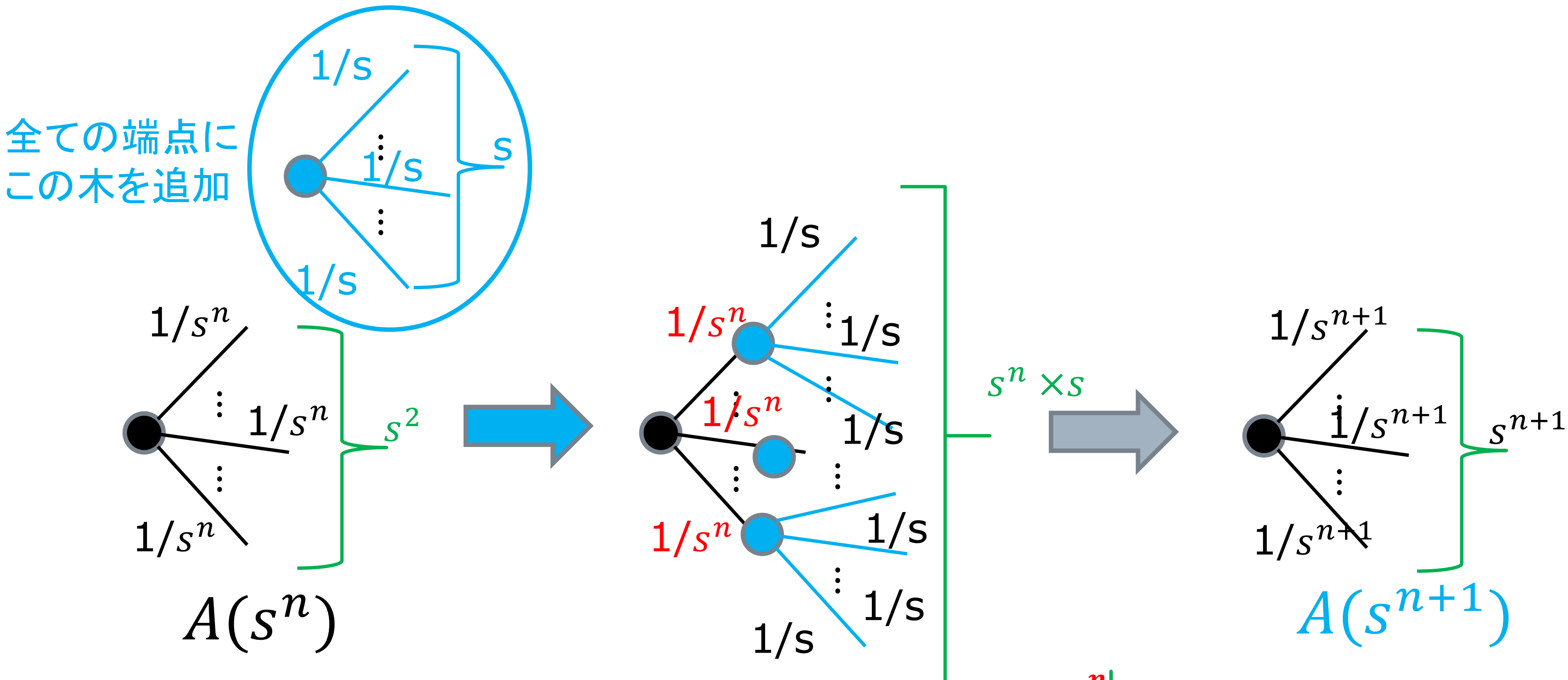
先の条件 3. で、シャノンのあげた例

H(1/2, 1/2)で表される選択と、H(2/3, 1/3)で表される選択を、次のように連続して行なったとする。



$$H(1/2, 1/3, 1/6) = H(1/2, 1/2) + 1/2\, H(2/3, 1/3)$$

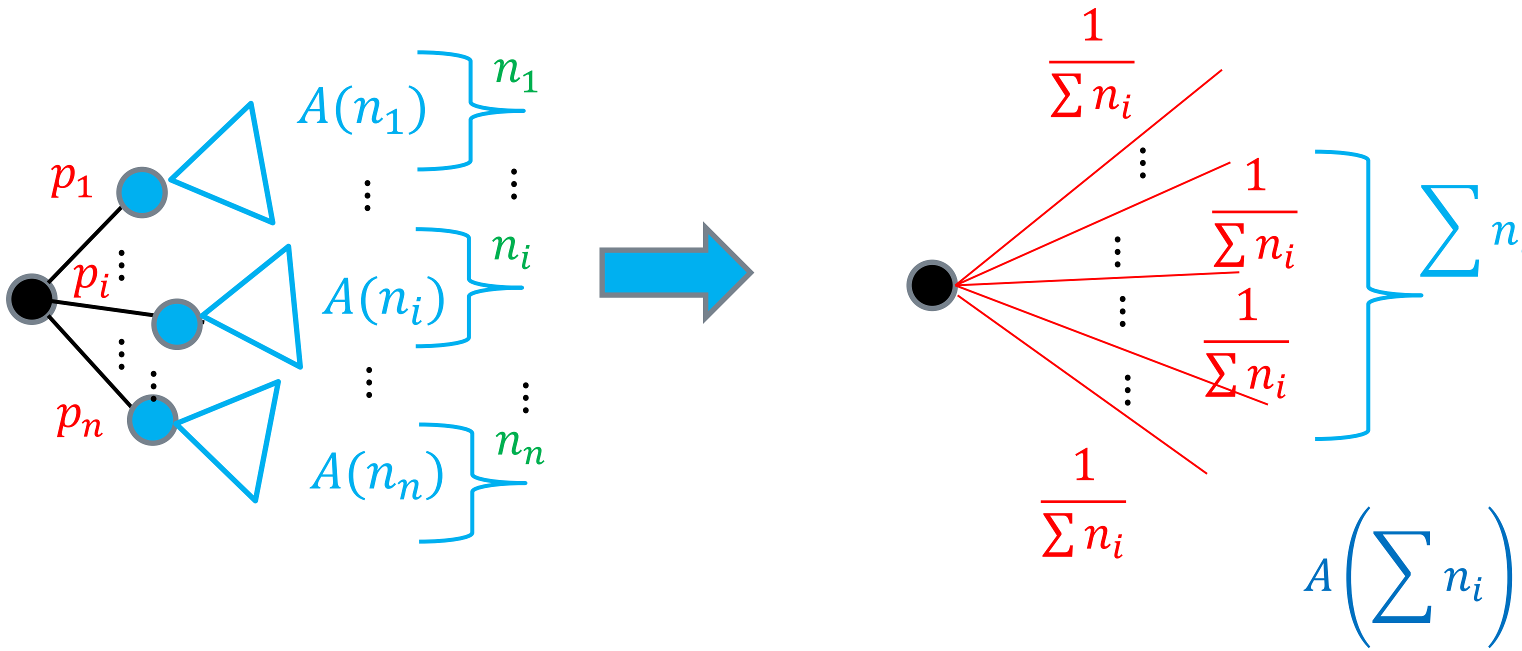
シャノンが行った $A(s^m) = mA(s)$ の導出



$$A(s^{n+1}) = A(s^n) + \overbrace{\frac{1}{s^n} A(s) + \frac{1}{s^n} A(s) + \cdots + \frac{1}{s^n} A(s)}^{s^n}$$

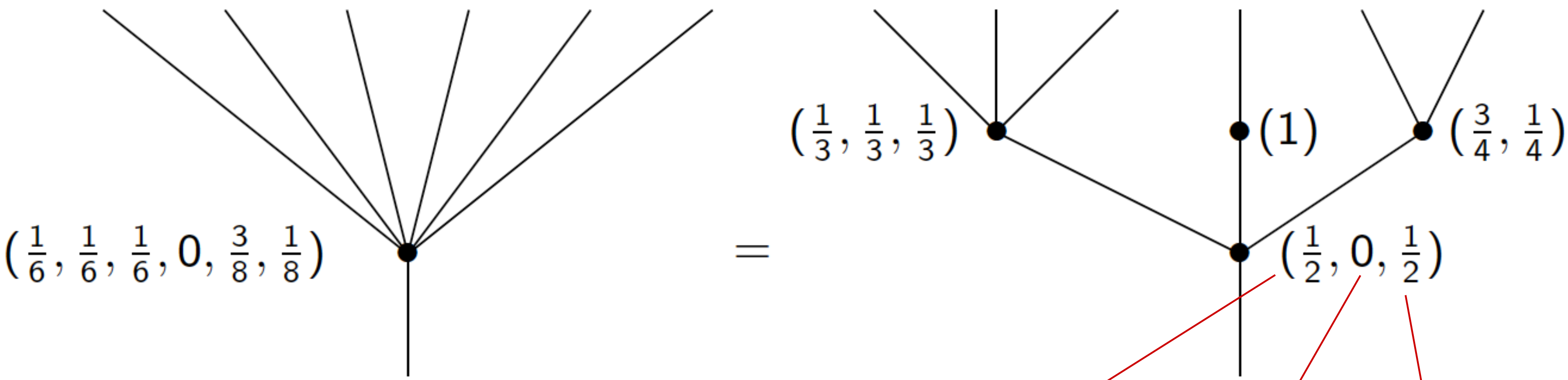
$$= nA(s) + A(s) = (n+1)A(s)$$

シャノンが行った $A\left(\sum n_i\right) = H(p_1 \cdots p_n) + \sum p_i A(n_i)$ の導出



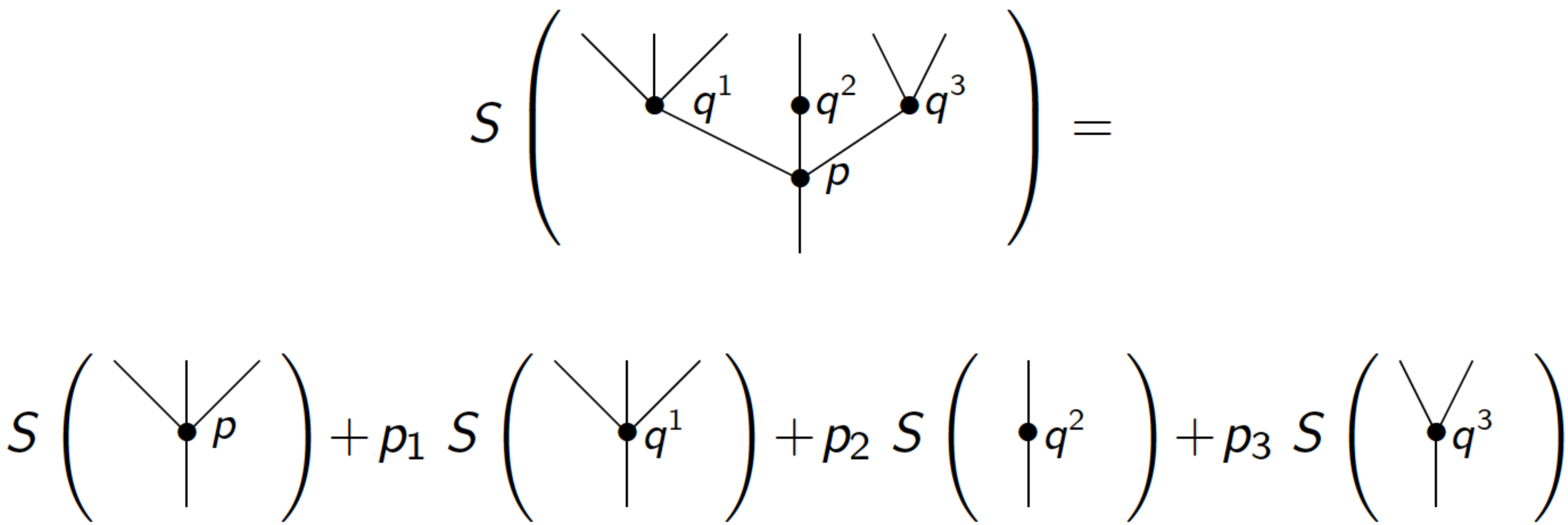
$$A\left(\sum n_i\right) = H(\mathrm{p}_1 \cdots\ p_n) + \mathrm{p}_1 \mathrm{A}(\mathrm{n}_1) + \mathrm{p}_2 \mathrm{A}(\mathrm{n}_2) + \cdots + p_n A(n_n)$$

図の向きを変えよう



$$S(\tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, 0, \tfrac{3}{8}, \tfrac{1}{8}) = S(\tfrac{1}{2}, 0, \tfrac{1}{2}) + \boxed{\tfrac{1}{2}} S(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}) + \boxed{0} S(1) + \boxed{\tfrac{1}{2}} S(\tfrac{3}{4}, \tfrac{1}{4})$$

Chain ルール



$$S\left(p \circ (q^1, q^2, \cdots, q^n)\right) = S(p) + \sum_{i=1}^{n} p_i S(q^i)$$

chain ルールの継承と発展

1948年　Shannon 「第三の仮定」

1956年　Faddeev 「Chain ルールの定式化」

2011年　Baez, Fritz, Leinster 「Entropy as a Functor」

2014年　Baez, Fritz「相対エントロピーへの応用」

2021年　Bradley 「Entropy as a Operad Derivation

Chainルールで
シャノン・エントロピーを特徴づける

シャノン・エントロピーの式を導く他のアプローチの探求

シャノンが、「三つの条件」から、シャノン・エントロピーHの式、

$$H(p_1, p_2, \ldots p_n) = -K \sum \mathrm{p_i} \log \mathrm{p_i} \quad (K\text{は}K > 0\text{なる定数})$$

を導いたことは、既に見てきた。

その後も、シャノン・エントロピーHの式を導く、数学的「条件」は何かという研究が行われた。

代表的なのは、Faddeevによる、次の論文である。

D. K. Faddeev, On the concept of entropy of a finite probabilistic scheme

https://arrowtheory.com/pub/notes/025-faddeev-entropy.html (*Arina Zinovyevaによる英訳*)

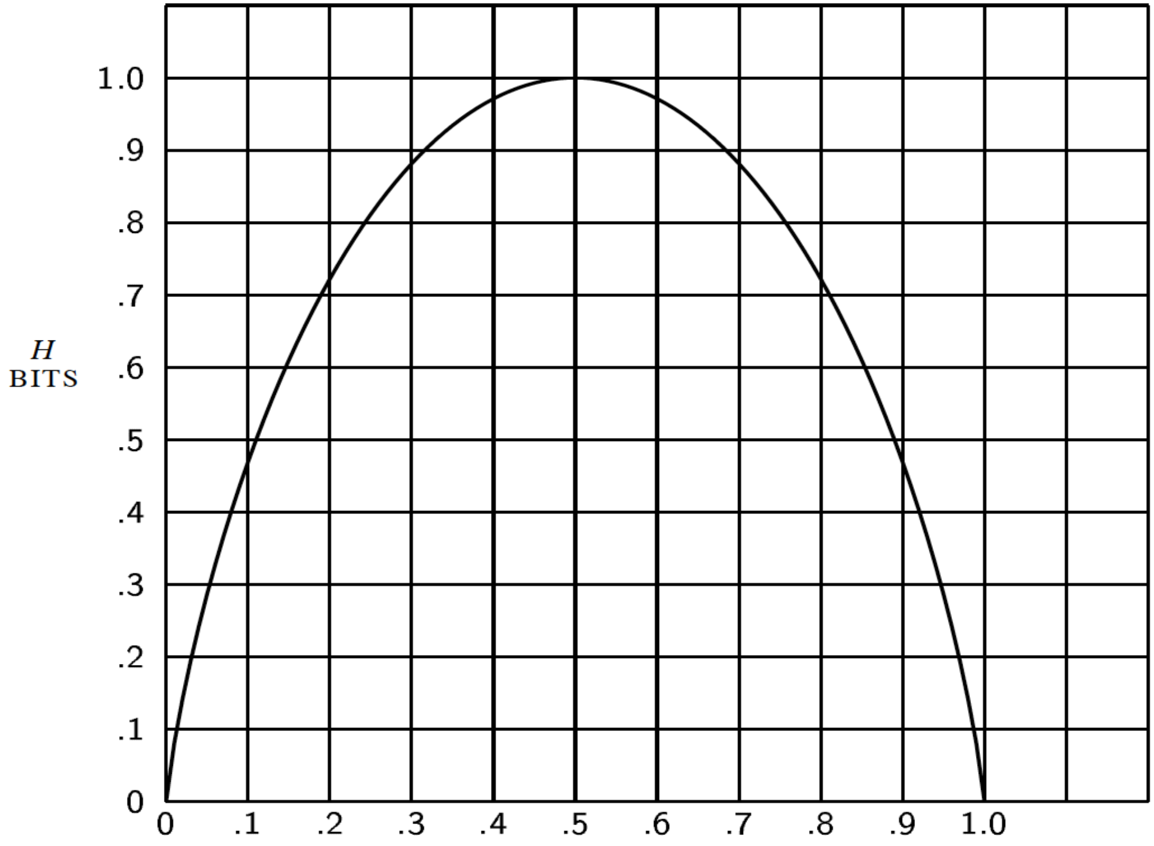
Faddeevが証明したこと

Faddeevは、次の三つの公理から、シャノン・エントロピーの式が導かれることを示した。

1. $0 \leq p \leq 1$の時、$H(p, 1-p)$は連続的で、少なくとも一つの点で正である。
2. $H(p_1, p_2, \ldots p_n)$は、$p_1, p_2, \ldots p_n$について、対称的な関数である。
3. $n \geq 2$の時、$p_n = q_1 + q_2$として、
$$H(p_1, p_2, \ldots p_{n-1}; q_1, q_2) = H(p_1, p_2, \ldots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$$

この公理を説明してみよう。

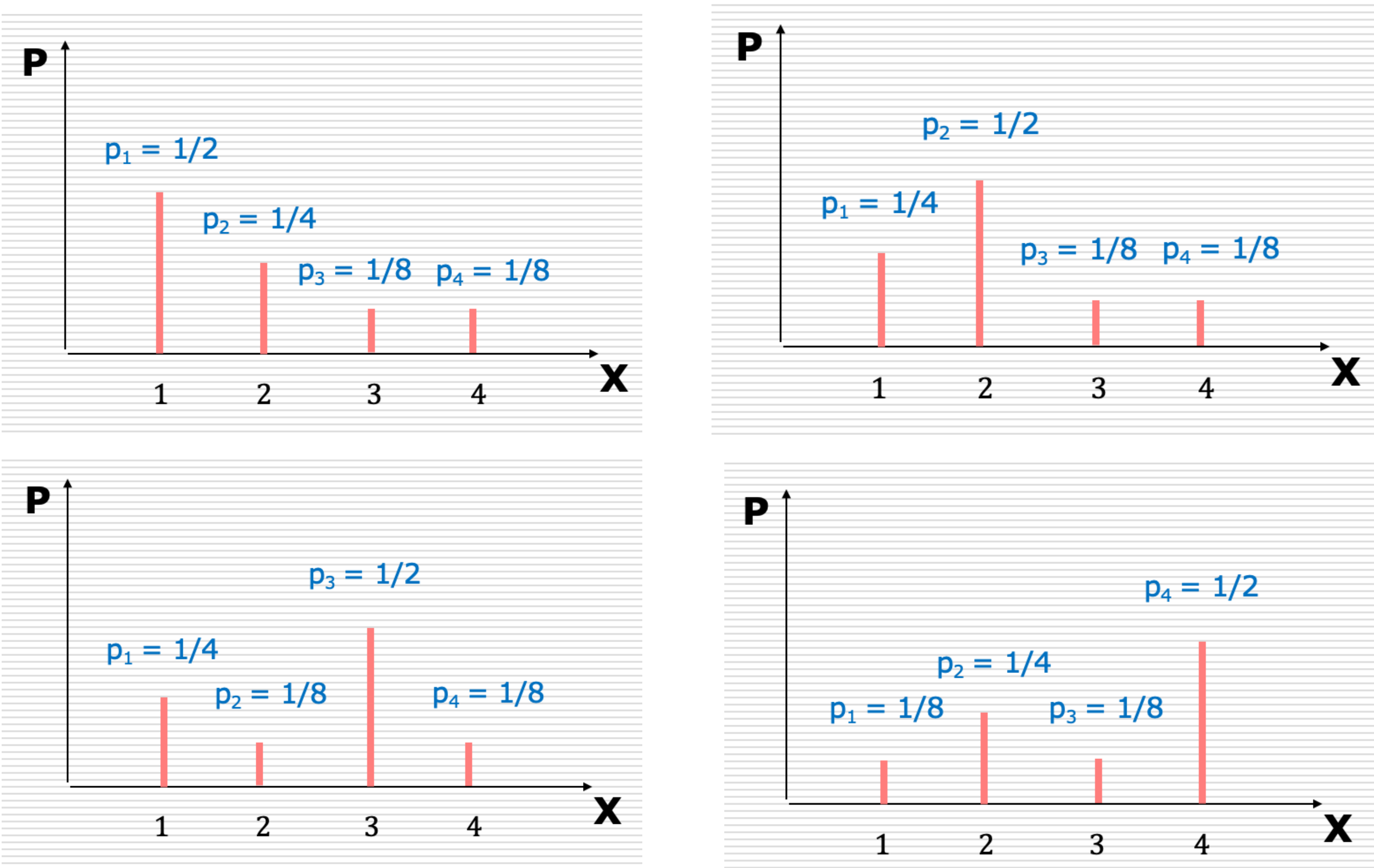
Faddeevの公理 1 のイメージ
H(p, 1 − p)の実際のグラフ



$p + q = 1$の時の

$H(\mathrm{p,q}) = -\mathrm{p}\log\mathrm{p} - \mathrm{q}\log\mathrm{q} = \mathrm{p}\log\left(\frac{1}{\mathrm{p}}\right) + \mathrm{q}\log(\frac{1}{\mathrm{q}})$ のグラフ

Faddeevの公理 2 のイメージ
確率変数を置換しても、エントロピーは変わらない



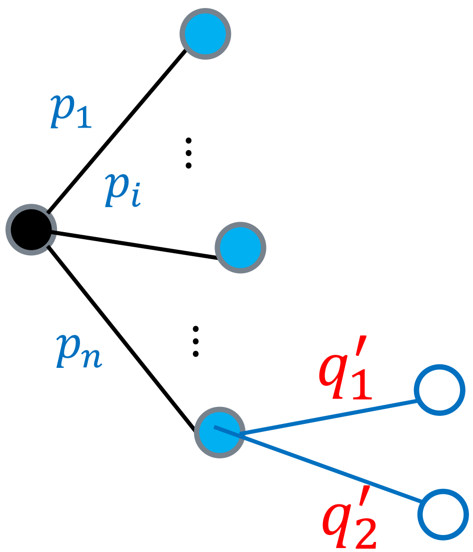
Faddeevの公理 3 のイメージ
単純な形のChain ルール

公理 3

$$n \geq 2, \qquad p_n = q_1 + q_2$$

$$H(p_1, p_2, \dots p_{n-1}; q_1, q_2) = H(p_1, p_2, \dots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$$

は、次のように図で表せる。



ただし、

$$q_1' = \frac{q_1}{p_n}, \quad q_2' = \frac{q_2}{p_n}$$

とおいた。

$q_1' + q_2' = \frac{q_1 + q_2}{p_n} = \frac{p_n}{p_n} = 1$ である。

Faddeevは、公理 3 から
基本的なChain ルールを定式化した

Lemma 6. From 2', 3', we conclude that:

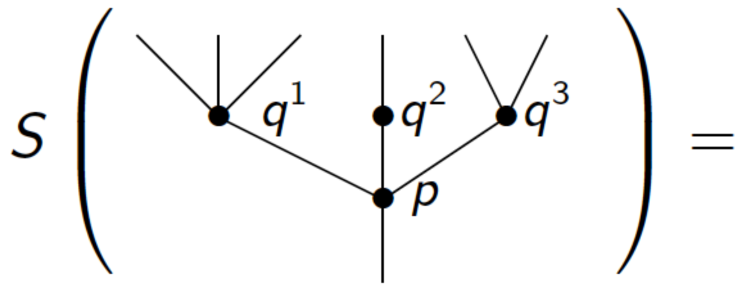
$$H(q_{11},\ldots,q_{1m_1};q_{21},\ldots,q_{2m_2};\ldots;q_{n1},\ldots,q_{nm_n}) = H(p_1,\ldots,p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i},\ldots,\frac{q_{im_i}}{p_i}\right).$$

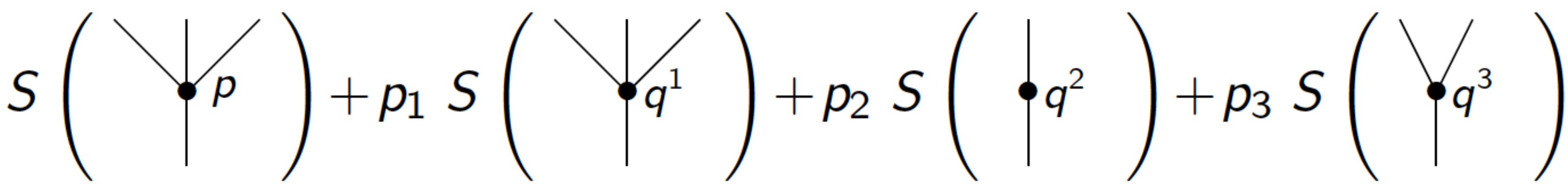
Here $p_i = q_{i1}+\ldots+q_{im_i}$.

この式は、次の基本的なChainルールに等しい。

$$S\big(p\circ(q^1,q^2,\cdots,q^n)\big) = S(p) + \sum_{i=1}^{n} p_i S(q^i)$$

Chain ルールのイメージ





$$S\left(p \circ (q^1, q^2, \cdots, q^n)\right) = S(p) + \sum_{i=1}^{n} p_i S(q^i)$$

Leinsterが証明したこと

Tom Leinster は、Faddeevの条件から出発して、さらに、次のことを証明した。

シャノン・エントロピーをHとし、$(I : \Delta_n \to \mathbb{R})_{n\geq 1}$を、確率分布を実数に写す関数列とする。この時、次の二つの条件は等しい。

i. 関数 Iは連続で、Chain ルールを満たす

$$I\big(\boldsymbol{w} \circ (\boldsymbol{p}^1, \ldots, \boldsymbol{p}^n)\big) = I(\boldsymbol{w}) + \sum_{i=1}^{n} w_i \, I(p^i)$$

ii. ある $c \in \mathbb{R}$ について、$I = cH$

Leinsterが証明したこと

別の言葉で言えば、シャノン・エントロピーは、連続性とChain・ルールによって、一意に特徴づけられるのである。

Tom Leinster, “Entropy and Diversity” Theorem 2.5.1

https://arxiv.org/pdf/2012.02113.pdf

シャノン・エントロピーは
Chainルールを満たすことを確かめる

シャノン・エントロピーは、確率分布の合成に関してChainルールを満たす

ある関数が、連続でかつchain ルールを満たすならば、その関数はシャノン・エントロピーに等しいと言えるのだが、その証明は、少し面倒である。

ただ、その逆、シャノン・エントロピーの定義から出発して、それがChain ルールを満たすことを示すのは容易である。（連続性は自明である）

ここでは、シャノン・エントロピーは、確率分布の合成に関してChainルールを満たすことを示してみよう。

$\partial(x) = -xlog(x)$なる$\partial(x)$を導入する

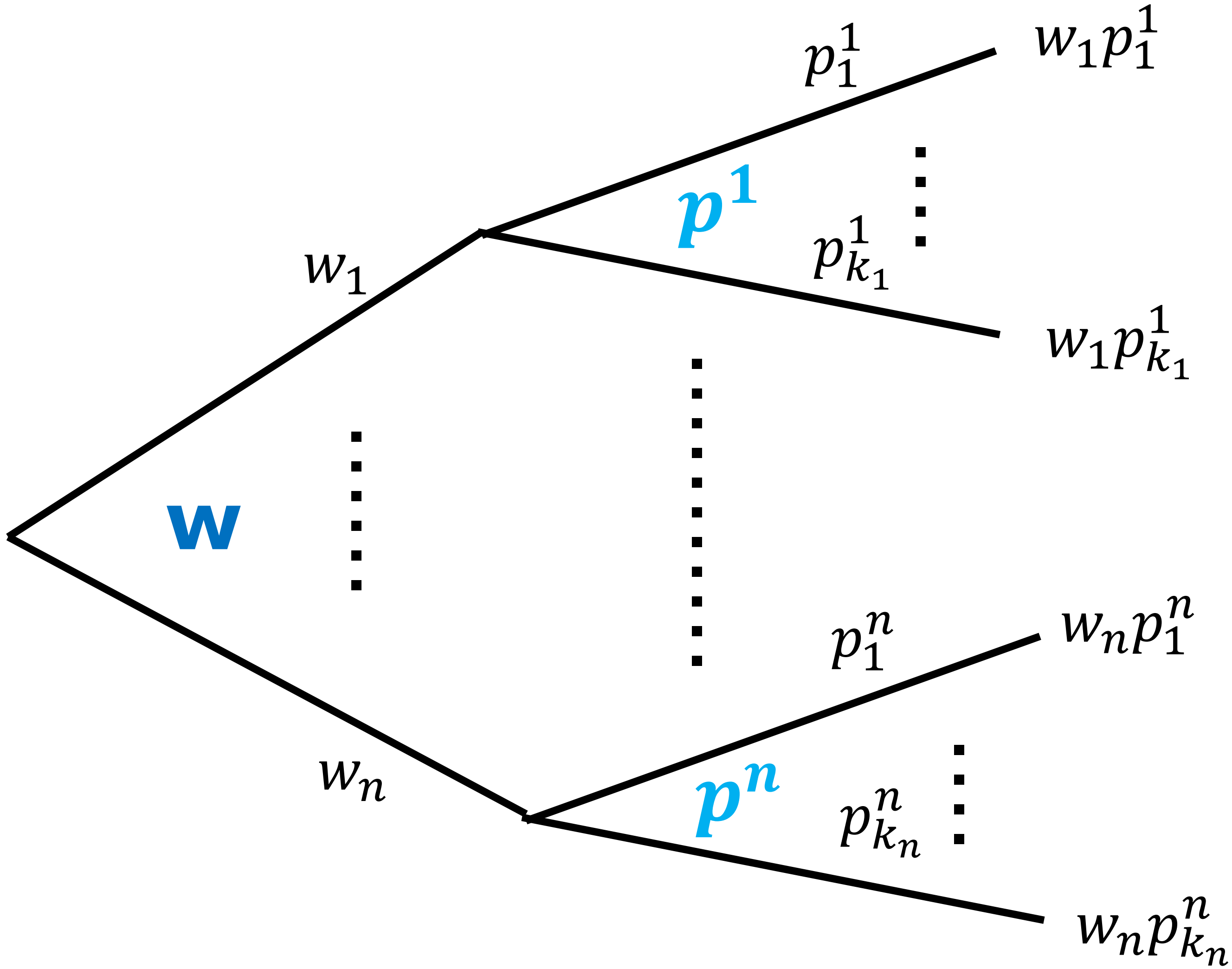
$x > 0$ の時、$\partial(x) = -xlog(x)$
$x = 0$の時、$\partial(x) = 0$ とする。

この時、確率分布 **p** のシャノン・エントロピー$H(\boldsymbol{p})$は、次のように表すことができる。

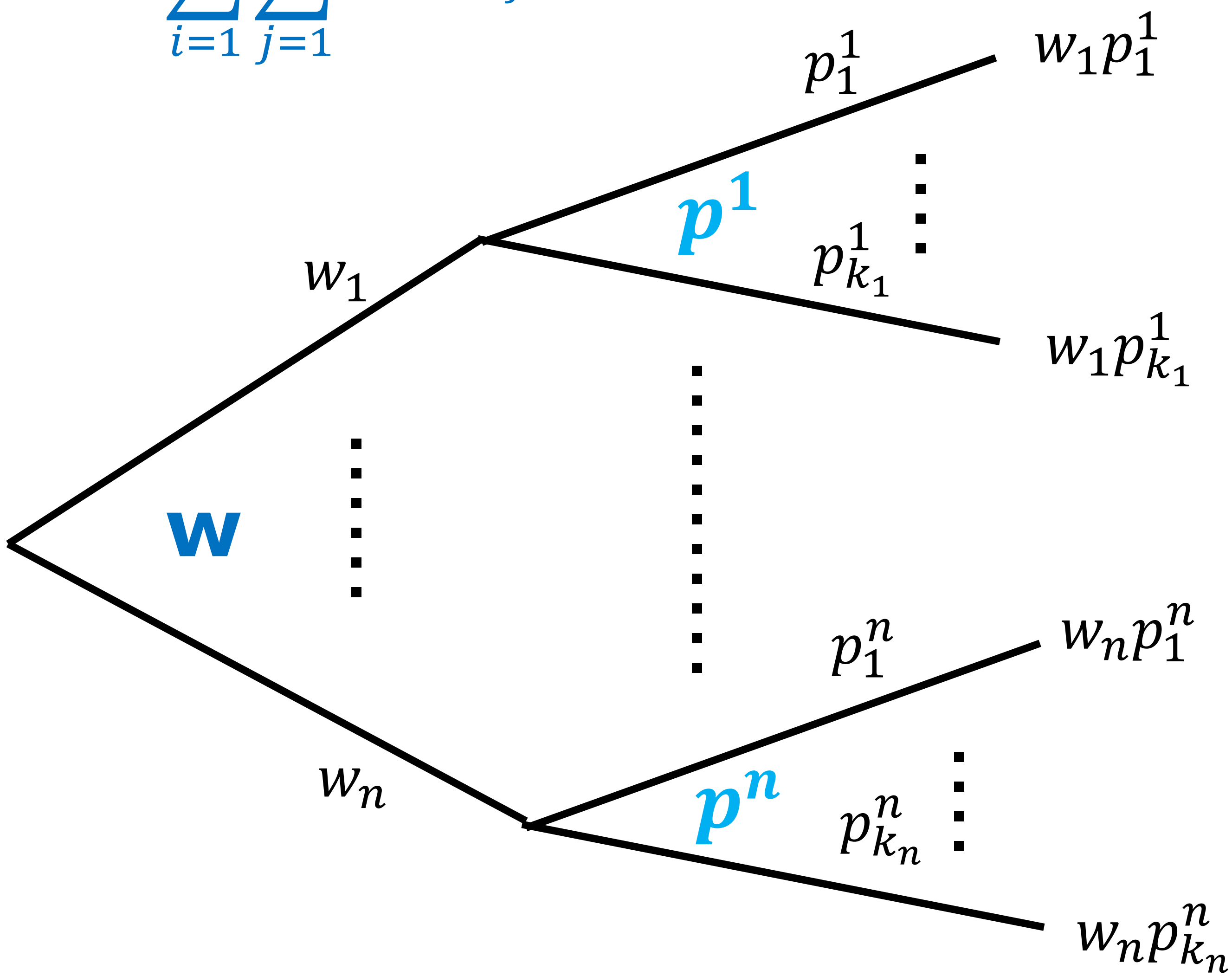
$$H(p) = \sum_{i=1}^{n} \partial(p_i)$$

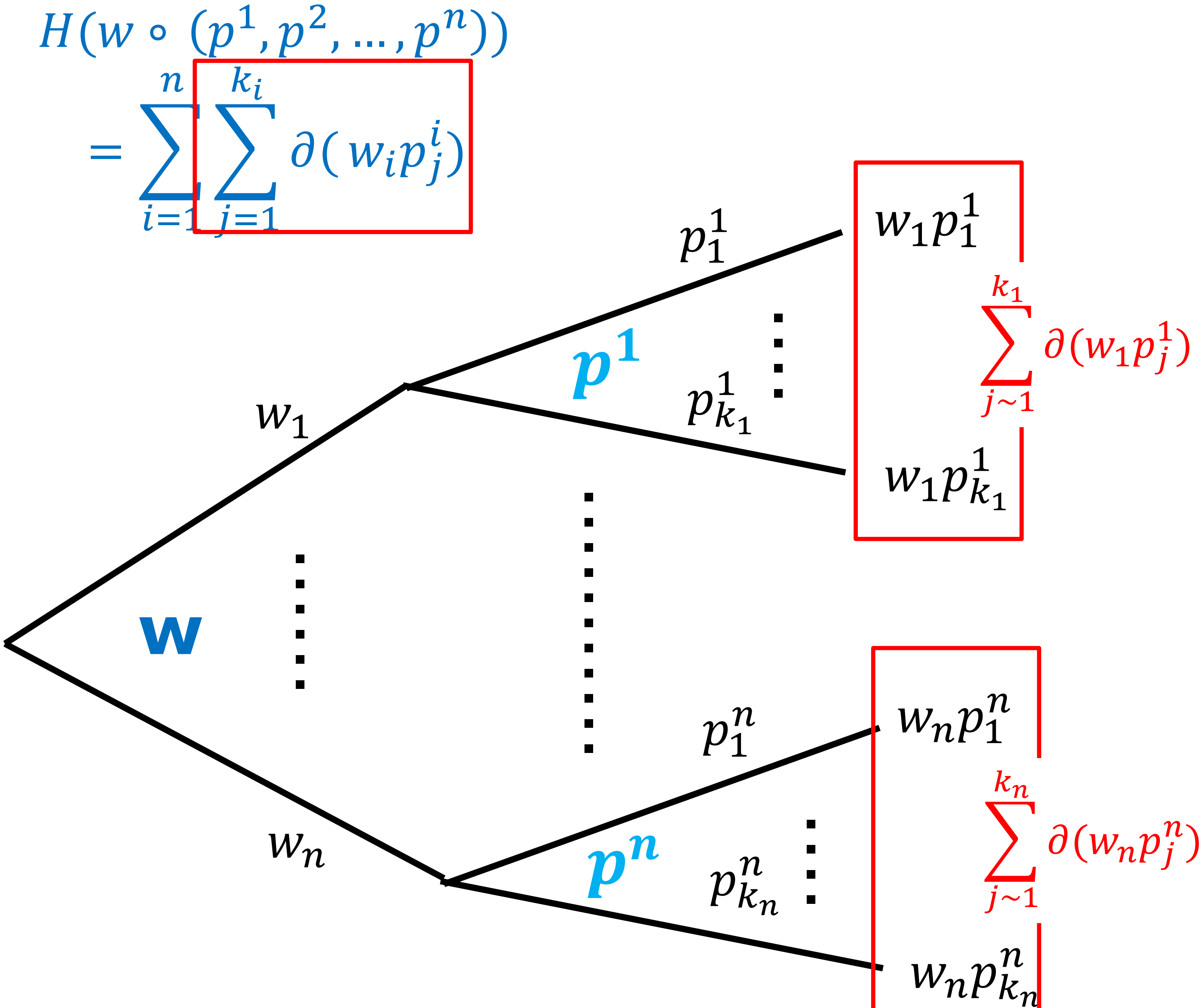
$$\partial(xy) = -xylog(xy) = -xylog(x) - xylog(y)$$
$$\partial(x)y + x\partial(y) = -xlog(x)y - xylog(y)$$
$$\therefore \partial(xy) = \partial(x)y + x\partial(y)$$

確率分布 $\boldsymbol{w}$と確率分布$(p^1, p^2, \ldots, p^n)$を
合成した確率分布 $w \circ (p^1, p^2, \ldots, p^n)$



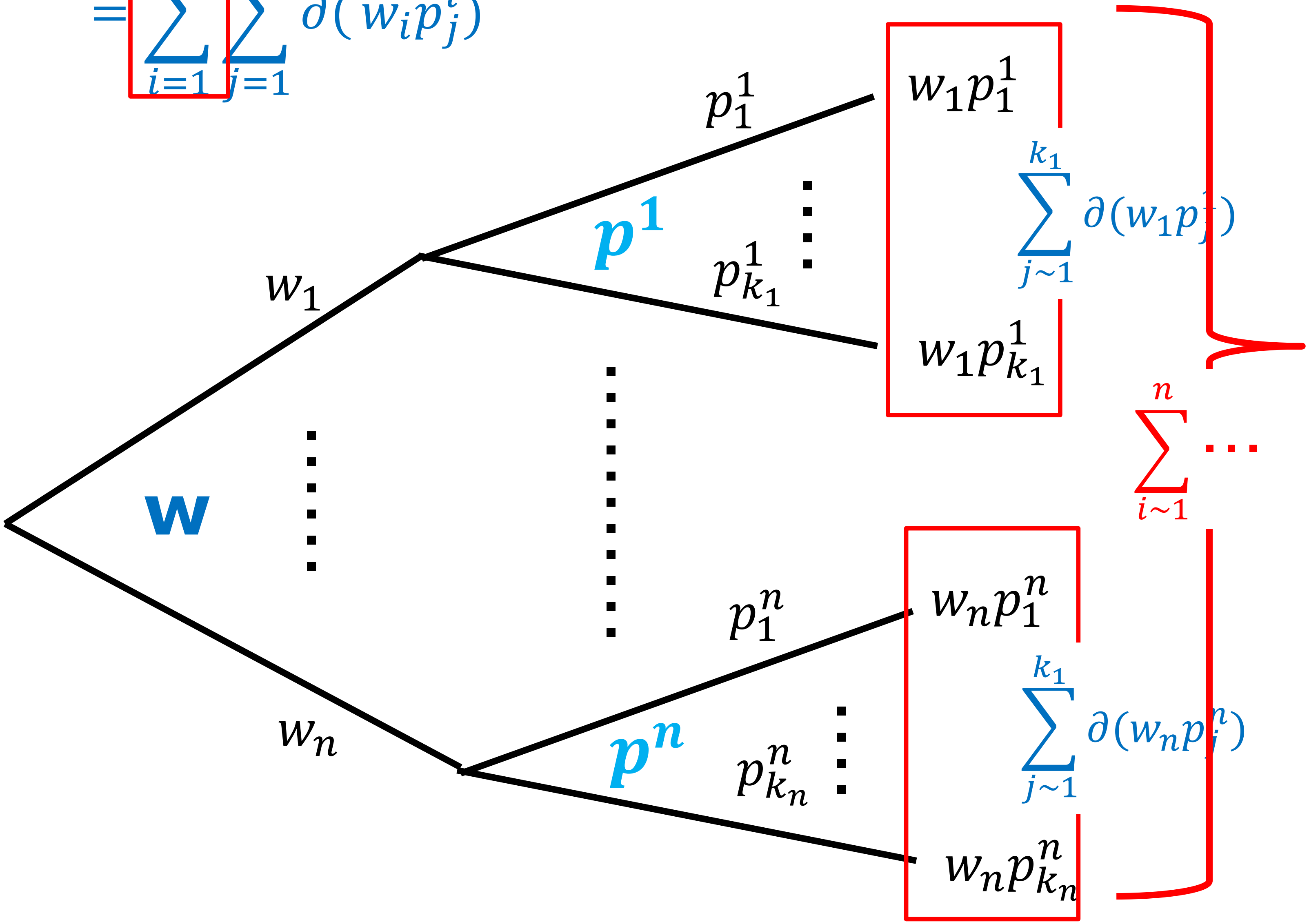
$$H(w \circ (p^1, p^2, \dots, p^n)) = \sum_{i=1}^{n} \sum_{j=1}^{k_i} \partial(w_i p_j^i)$$





$H(w \circ (p^1, p^2, \dots, p^n))$
$= \sum_{i=1}^{n} \boxed{\sum_{j=1}^{k_i} \partial(w_i p_j^i)}$
w
w_1
w_n
p^1
p^n
p_1^1
$p_{k_1}^1$
p_1^n
$p_{k_n}^n$
$w_1 p_1^1$
$w_1 p_{k_1}^1$
$w_n p_1^n$
$w_n p_{k_n}^n$
$\sum_{j\sim 1}^{k_1} \partial(w_1 p_j^1)$
$\sum_{j\sim 1}^{k_n} \partial(w_n p_j^n)$

$$H(w \circ (p^1, p^2, \ldots, p^n)) = \boxed{\sum_{i=1}^{n}} \sum_{j=1}^{k_i} \partial(w_i p_j^i)$$



$$H(w \circ (p^1, p^2, \dots, p^n))$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{k_i} \partial(w_i p_j^i)$$

$$\partial(xy) = \partial(x)y + x\partial(y)$$

$$= \sum_i \sum_j (\partial(w_i)\, p_j^i + w_i \partial(p_j^i))$$

$$\sum_j p_j^i = 1$$

$$= \sum_i \partial(w_i) + \sum_i w_i \sum_j \partial(p_j^i)$$

$$H(w) = \sum_i \partial(w_i), \quad H(p^i) = \sum_j \partial(p_j^i)$$

$$= H(\boldsymbol{w}) + \sum_i w_i H(\boldsymbol{p^i})$$

$$H(w \circ (p^1, p^2, \dots, p^n)) = H(\boldsymbol{w}) + \sum_i w_i H(\boldsymbol{p^i})$$

第四部

Baez：新しいエントロピー論の登場
Entropy as a Functor

第四部 Agenda
Baez：新しいエントロピー論の登場
Entropy as a Functor

- 「情報の損失」としてのエントロピー
- 測度を保存する関数と
  確率分布の表現Δ_nについて
- 「情報の損失」の性質を考える
- 「情報の損失」の性質を公理化し、
  シャノン・エントロピーの式を導く
- Entropy as a Functor

バエズが考えたこと　1

「情報の損失」としてのエントロピー

バエズによる
「情報の損失」としてのエントロピー論

この章では、現代のエントロピー論の飛躍の突破口となった、2011年のバエズらの論文 “A Characterization of Entropy in Terms of Information Loss”
https://arxiv.org/pdf/1106.1791.pdf を見ていこうと思う。

論文のタイトルが示すように、「情報の損失」としてエントロピーを特徴づけようというものだ。

バエズ自身による論文の概要の紹介し、続いてそれを補足する。
“A Characterization of Entropy”
https://johncarlosbaez.wordpress.com/2011/06/02/a-characterization-of-entropy/

エントロピーの変化に注目する

$$S(p) = -\sum_{i \in X} p_i \ln p_i$$

この奇妙な見かけの式は、多くの仕方で正当化できる。

我々の新しい方法は、エントロピーそのものではなく、エントロピーの変化に注目しようということと関わっている。

このことは、多くの理由で意味を持つ。

例えば、物理では通常はエントロピーを直接には測定しない。そのかわりにエントロピーの変化を次のような経験的な事実を利用して測定する。すなわち、温度Tにある系が小さな量の熱 ΔQ を可逆なスタイルで吸収する時、エントロピーの変化は ΔQ/T である。

しかし、我々がエントロピーの変化に注目する本当の理由は、それがある本当に巧妙な定理を与えるからである。

確率測度を保存する関数

確率測度を持つ有限集合が二つあるとする。それを(X, p), (Y, q)としよう。

この時、関数 $f: (X, p) \to (Y, q)$が測度を保存するということを次のように定義する。

Y の任意の点 j上の確率q_jは、$f(i) = j$ であるXの点 i 上の確率p_i の和に等しい。

$$q_j = \sum_{i:\, f(i)=j} p_i$$

確率測度を保存する関数は
エントロピーを減少させる

この種の関数は、あるランダムな状況を他のランダムな状況に移す決定論的なプロセスである。

例えば、ある確率分布で選ばれた-10と10の間のランダムな整数があったとする。この数を二乗すると0と100の間のランダムな整数ができる。この種のプロセスは常にエントロピーを減少させる。

すなわち、与えられた測度を保存する関数 $f:(X,p)\to(Y,q)$に対して、次が言える。

$$f:(X,p)\to(Y,q)$$
$$S(p)\ \geq\ S(q)$$

決定論的なデータ処理は、情報を増やさない
情報を減らすだけである

熱力学の第二法則は、エントロピーは常に増大するというのだから、この式は直観に反していて、矛盾しているようにさえ見える！

しかし、ここには矛盾はない。もし、エントロピーを情報だと考えて、関数 f は、追加的にどんなランダムさも導入しないある種のデータ処理だと考えれば、この例は、もっと直観的な意味を持つ。

このようなプロセスは、情報の量を減らすだけである。例えば、数 −5 を二乗すれば、数 5を二乗したのと同じ値を得る。だから、わたしが「この数の二乗は 25 である」と君に告げたとすれば、私が「この数は−5 である」と君に告げるより、君により少ない情報を与えたことになる。

確率測度を保存する関数に対して、「情報の損失」を表す関数を定義する

こうした理由で、差 $F(p) - F(q)$を、関数 $f: (X, p) \to (Y, q)$の「情報の損失」と呼ぶ。

最初に、関数 $f: (X, p) \to (Y, q)$ を、$f: p \to q$ と短く表すことにしよう。Fをこのような任意の関数fに数$F(f) \in [0. \infty)$ を割り当てる関数だとしよう。

$$f: p \to q$$
$$F(f) = F(p) - F(q)$$

それを、我々は「情報の損失」と考える。

ここでは、シャノンのエントロピーを「情報の損失」という言葉で特徴づけようと思う。

バエズが考えたこと 1 補足

測度を保存する関数と
確率分布の表現 Δ_n について

確率測度を保存する関数

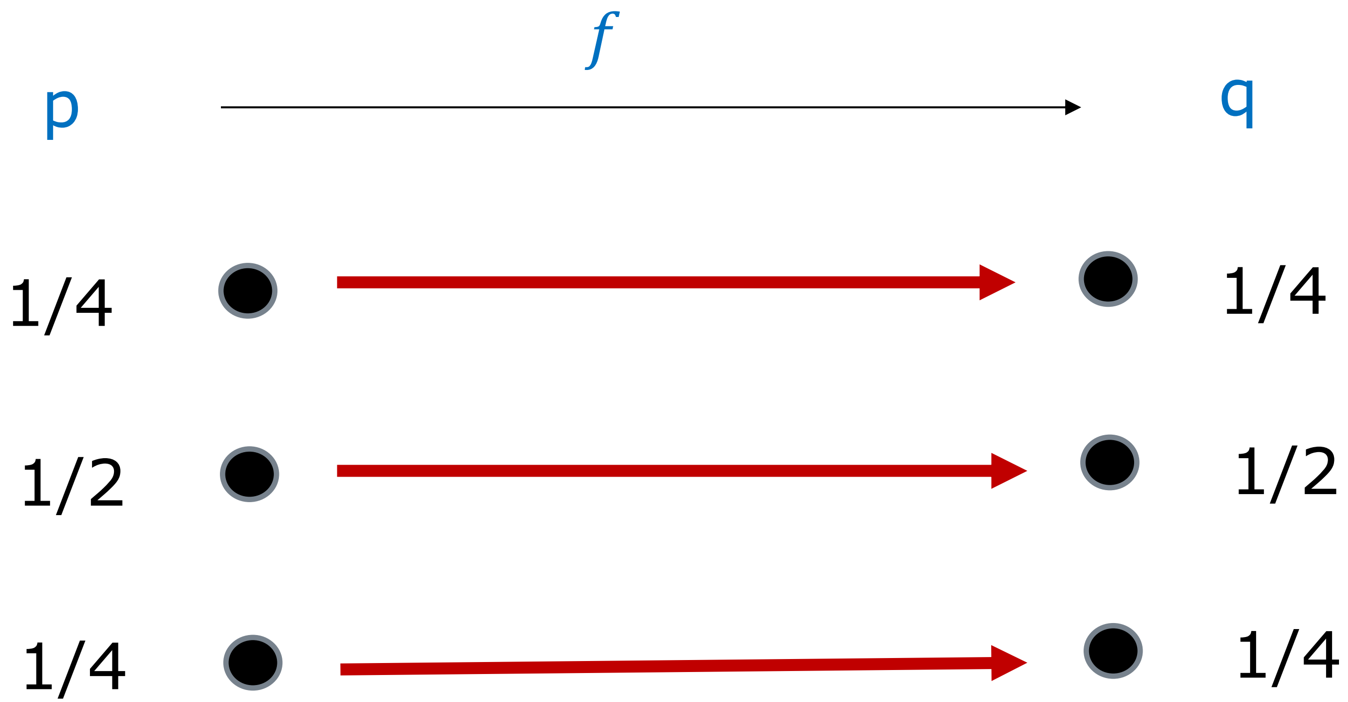
確率測度を持つ有限集合が二つあるとする。それを(X, p), (Y, q)としよう。

この時、関数 $f: (X, p) \to (Y, q)$が測度を保存するということを次のように定義する。

Y の任意の点 j上の確率q_jは、$f(i) = j$ であるXの点 i 上の確率 p_i の和に等しい。

$$q_j = \sum_{i:\, f(i)=j} p_i$$

測度を保存する関数の例　1

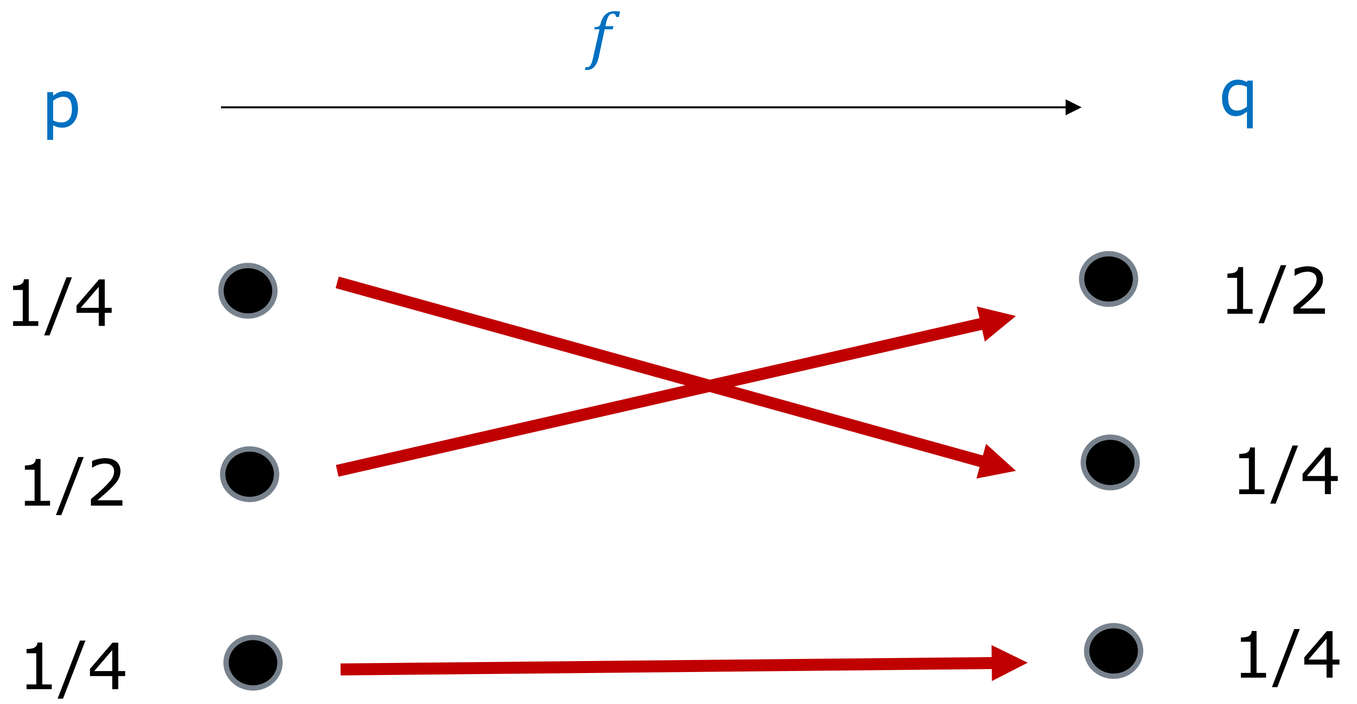


$S_p \approx 1.04$

$S_q \approx 1.04$

$$Loss(f) = S(p) - S(q) = 1.04 - 1.04 = 0$$

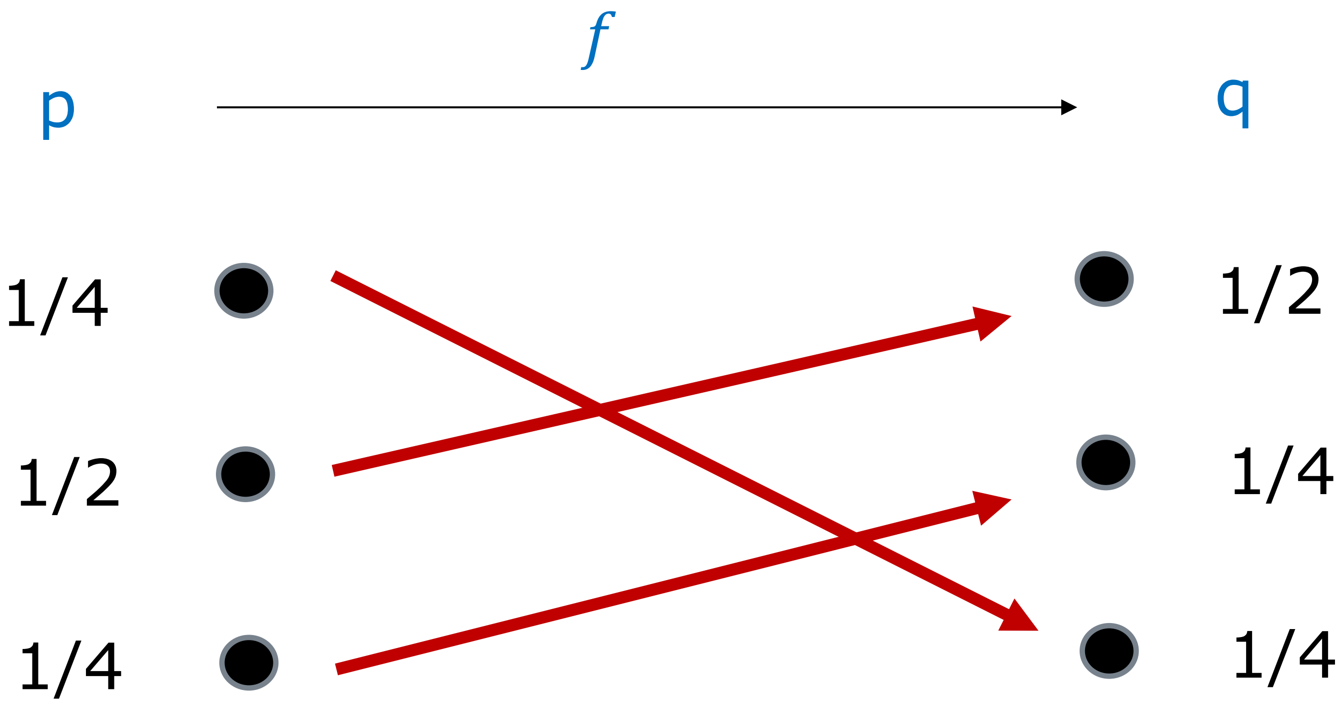
測度を保存する関数の例　2



$S_p \approx 1.04$　　　　$S_q \approx 1.04$

$$Loss(f) = S(p) - S(q) = 1.04 - 1.04 = 0$$

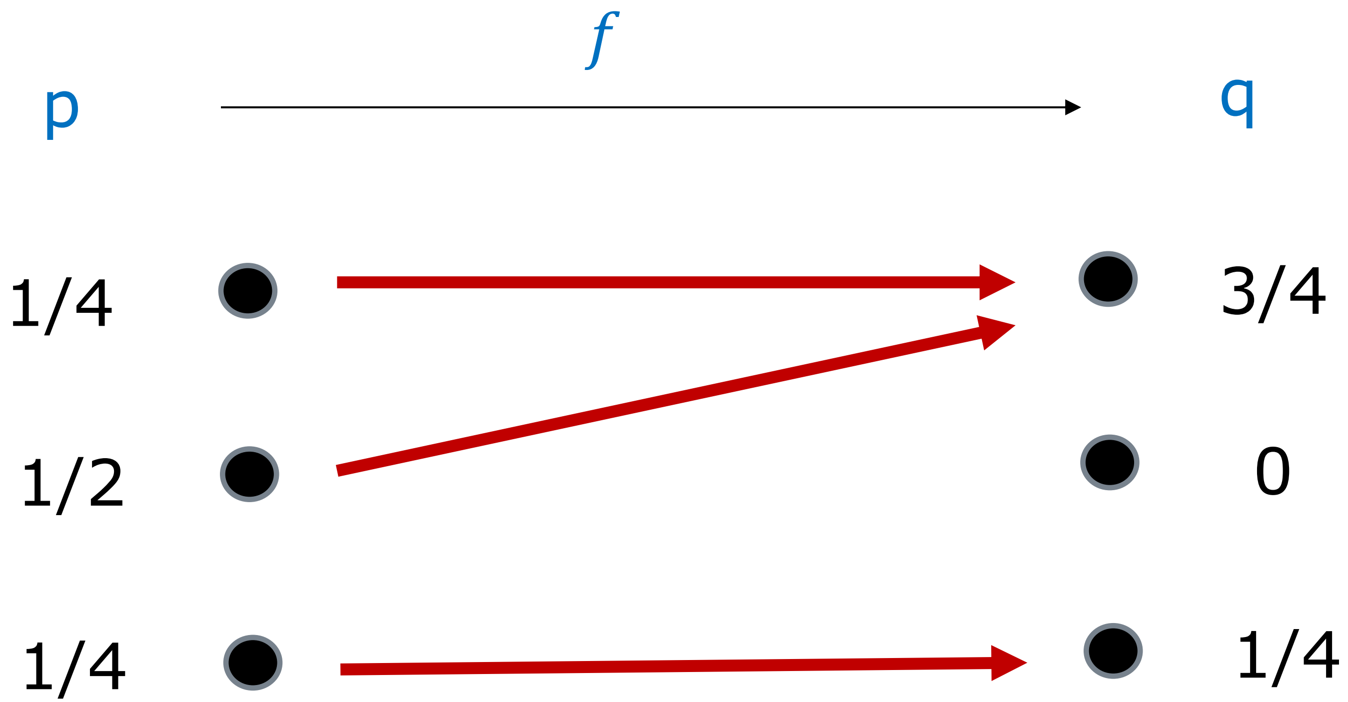
測度を保存する関数の例　3



$S_p \approx 1.04$　　　$S_q \approx 1.04$

$$Loss(f) = S(p) - S(q) = 1.04 - 1.04 = 0$$

測度を保存する関数の例　4



$$q_j = \sum_{i:\, f(i)=j} p_i$$

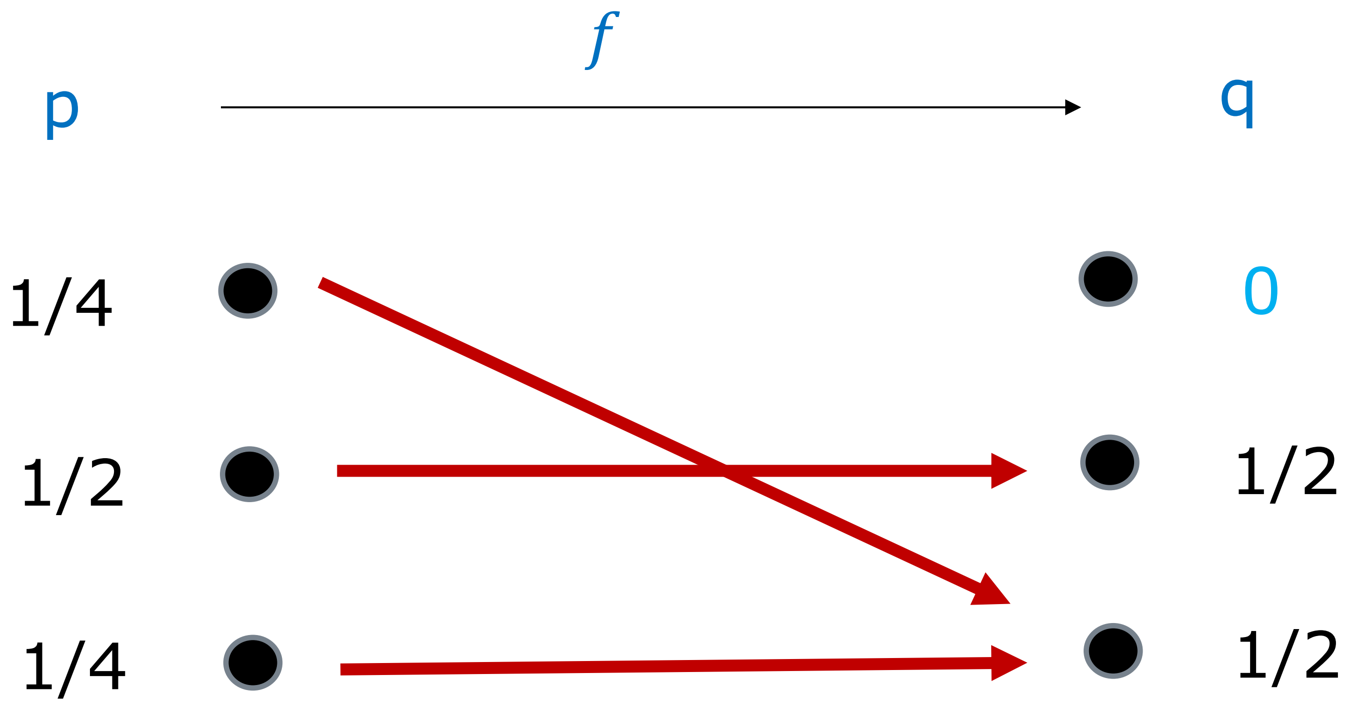
$$= 1/4 + 1/2$$

$S_p \approx 1.04$

$S_q \approx 0.56$

$$Loss(f) = S(p) - S(q) = 1.04 - 0.56 = 0.48$$

測度を保存する関数の例　5



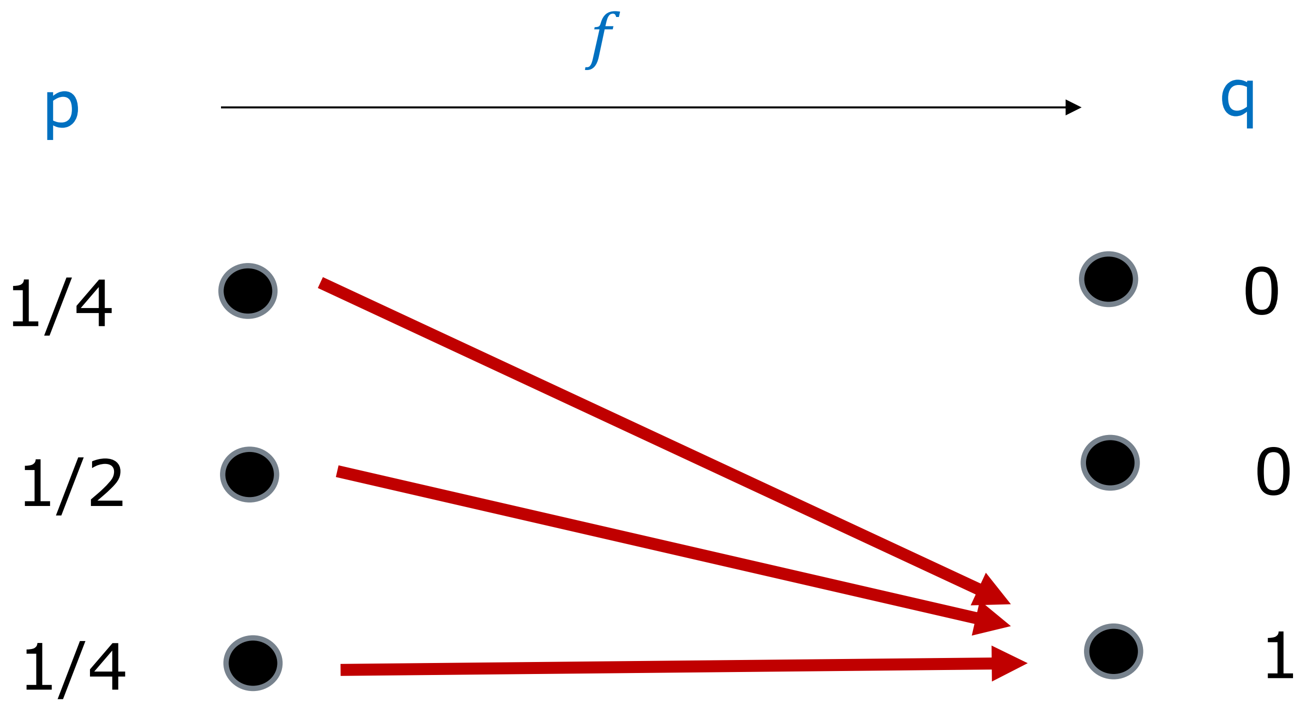
$$q_j = \sum_{i:\, f(i)=j} p_i$$

$$= 1/4 + 1/4$$

$S_p \approx 1.04$　　$S_q \approx 1$

$$Loss(f) = S(p) - S(q) = 1.04 - 1 = 0.4$$

測度を保存する関数の例　6



$$q_j = \sum_{i:\, f(i)=j} p_i$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{4}$$

$S_p \approx 1.04$　　$S_q \approx 0$

$$Loss(f) = S(p) - S(q) = 1.04 - 0 = 1.04$$

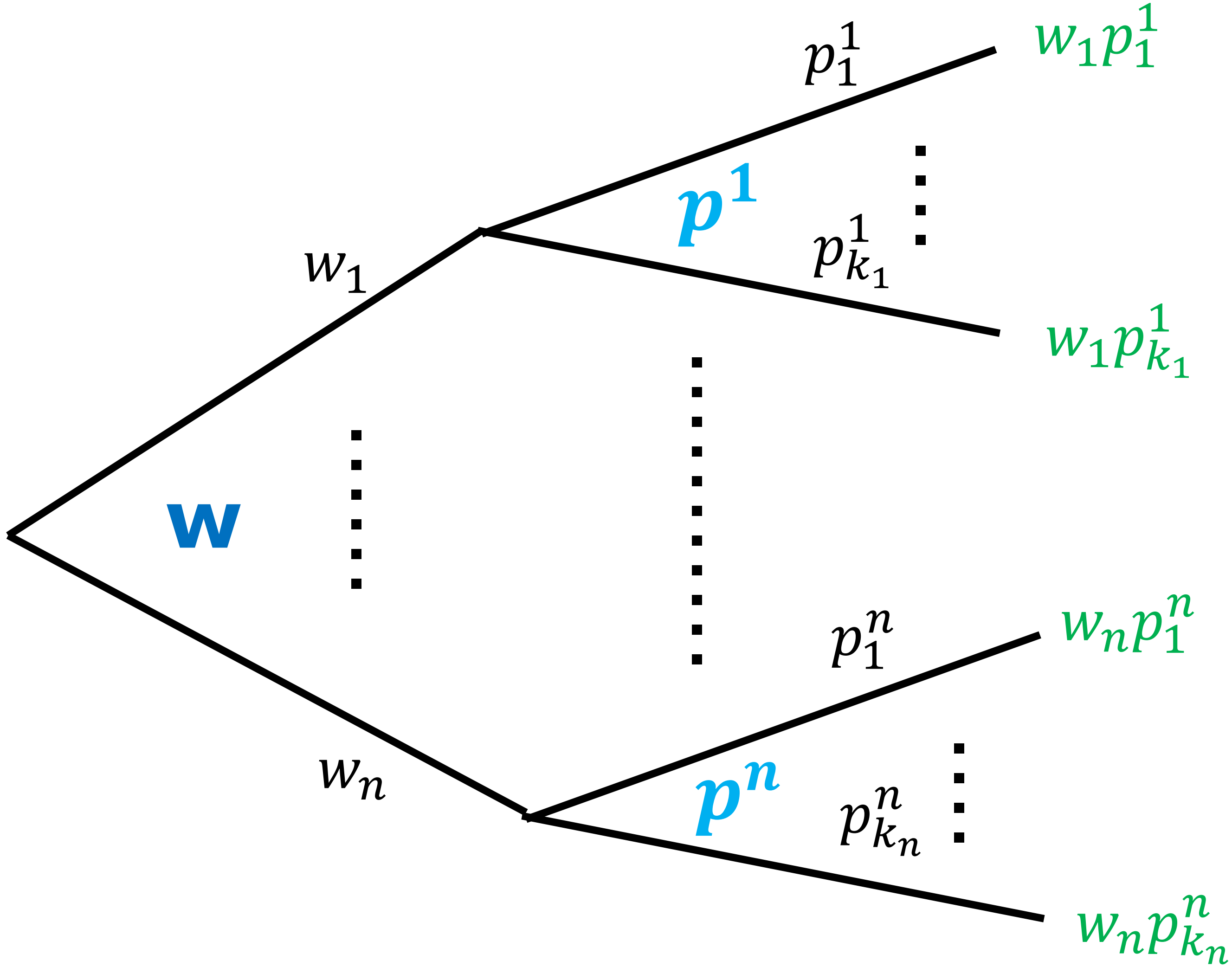
確率分布の表現 Δ_n について

フェアなコインがあったとする。そのコイントスで与えられる確率分布(X, p)を考える。X(表、裏）と考えても、X(勝、負)と考えても、X(H,T)と考えても、分布の形 p(1/2, 1/2) に変化はない。

基本的には、n個の確率変数からなる確率分布は、$\{1,2, \ldots, n\}$上の確率分布と考えることができる。確率変数の名前は、$\{1,2, \ldots, n\}$に、一対一に割り当てればいい。

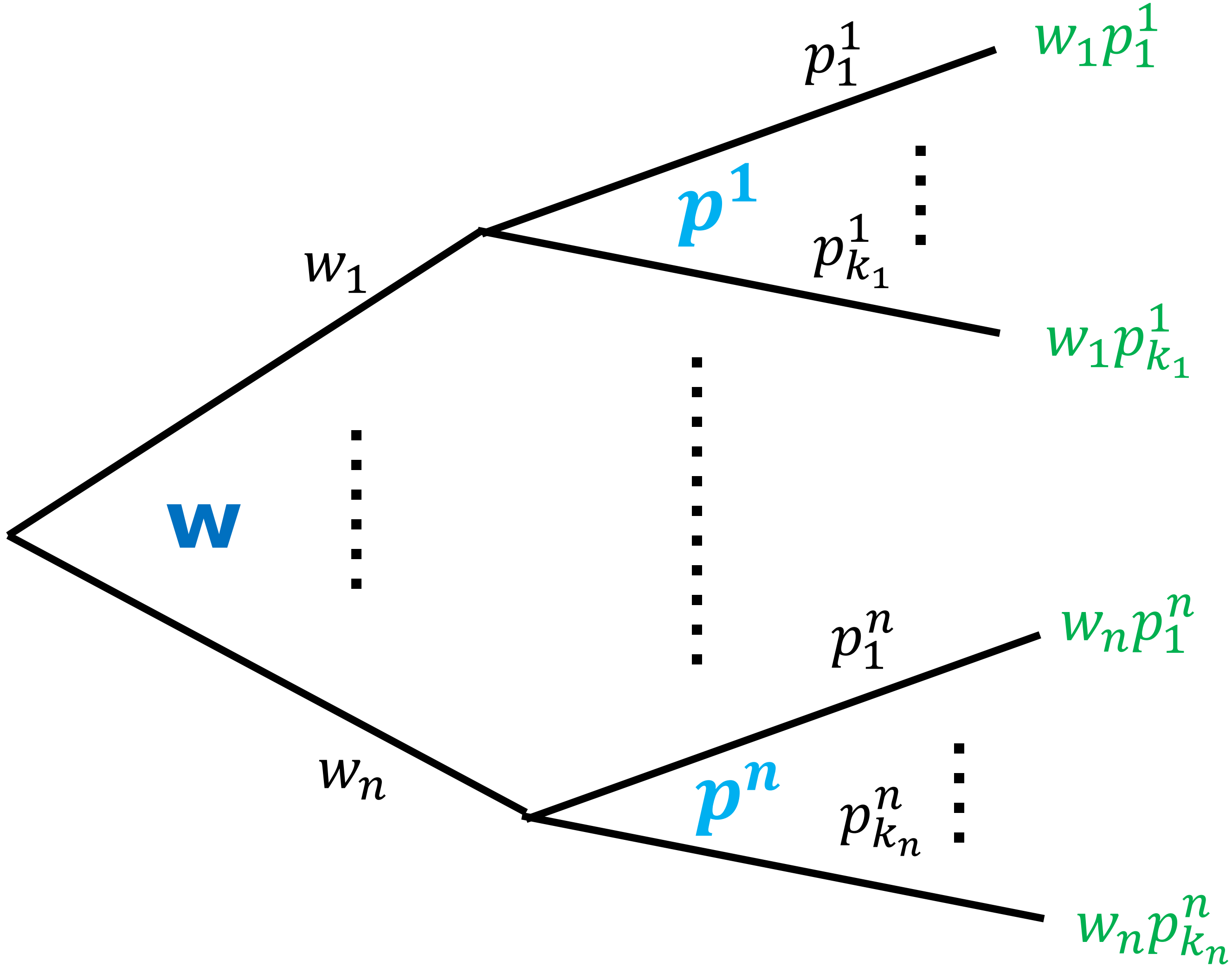
Δ_nで $\{1,2, \ldots, n\}$上の確率分布を表す。

確率分布$\boldsymbol{w}$と確率分布$(p^1, p^2, \ldots, p^n)$を合成した
確率分布 $w \circ (p^1, p^2, \ldots, p^n)$でこの表記を利用してみる



$$\boldsymbol{w} \in \Delta_n, \qquad \boldsymbol{p}^1 \in \Delta_{k_1}, \ldots, \boldsymbol{p}^n \in \Delta_{k_n}$$

$$(w_1p_1^1, \ldots, w_1p_{k_1}^1, \ldots, w_np_1^n, \ldots, w_np_{k_n}^n) \in \Delta_{k_1+k_2+\cdots k_n}$$

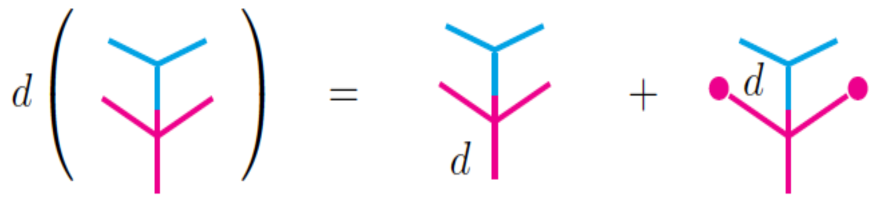


トポロジカルな simplex Δ^nの定義

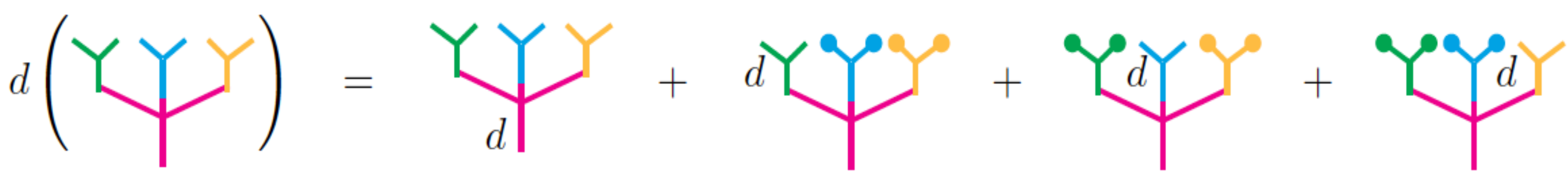
$$\Delta^n := \{(p_0, p_1, \ldots, p_n) \in \mathbb{R}^{n+1} \mid 0 \leq p_i \leq 1 \text{ and } \sum_{i=0}^{n} p_i = 1\},$$

$$\Delta^{n-1} = \Delta_n$$

Bradleyの“Entropy as a Topological Operad Derivation” から



$$d(p \circ_i q) = dp \circ_i q + p \circ_i dq$$



$$\boxed{d(p \circ (q^1, \dots, q^n))(x) = dp(\langle q^1, \mathbf{x}_1 \rangle, \cdots, \langle q^n, \mathbf{x}_n \rangle) + \sum_{i=1}^{n} p_i dq^i(\mathbf{x}_i).}$$

バエズが考えたこと　２

「情報の損失」の性質を考える

バエズが行おうとしたこと

- シャノンは、「三つの条件」から、シャノン・エントロピーの式を導いた。(本セミナーの第二部)
- Faddeev と Leinster は、「連続でchainルールに従う」という条件から、シャノン・エントロピーの式を導いた。(本セミナーの第三部)

バエズは、情報プロセスでの「情報の損失」という特徴から、シャノン・エントロピーの式を導びこうとする。

まず、「情報の損失」がどういう性質を持つのか見ていこう。

情報の損失 $Loss(f)$ の定義

測度を保存する関数 f があって、確率分布(X, p)を確率分布 (Y, q) に変えるとする。

$$f: (X, p) \longrightarrow (Y, q)$$

この時、情報の損失 $Loss(f)$ は、次のように定義される。

$$Loss(f) = S(p) - S(q)$$

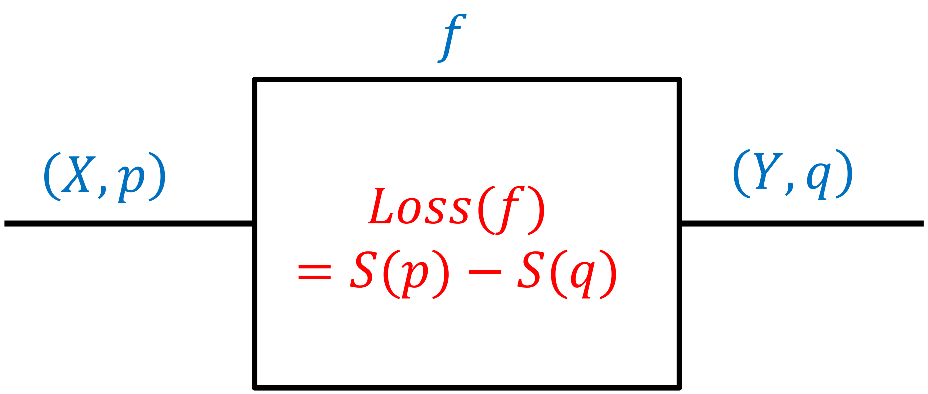
$S(p), S(q)$ は、確率分布 p, q のシャノン・エントロピー

測度を保存する関数 f は、エントロピーを減少させるので、

$Loss(f) = S(p) - S(q) \geq 0$である。

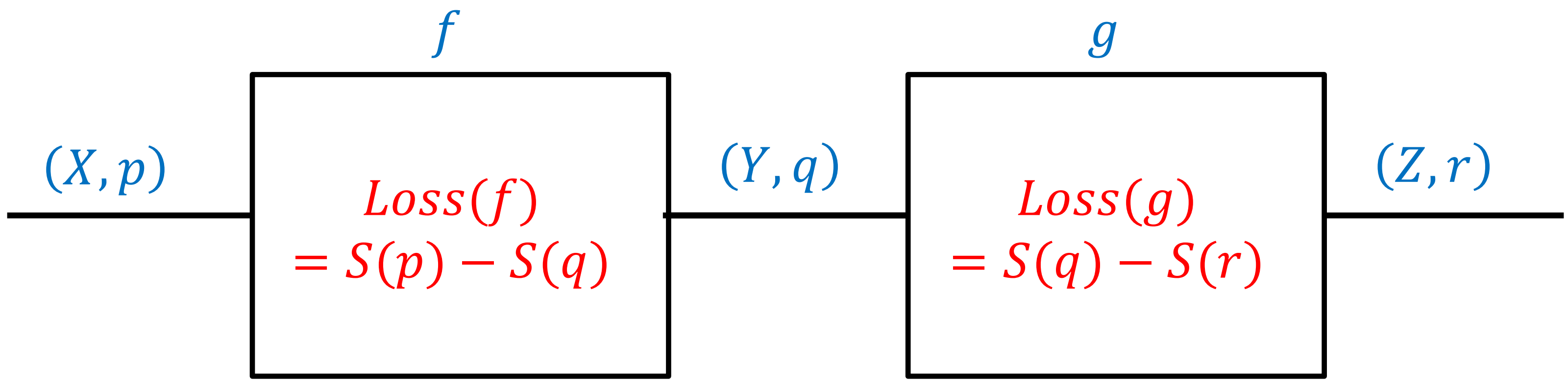
「情報の損失」をプロセスとして見る

情報の損失は、次のようなプロセスと考えることができる。

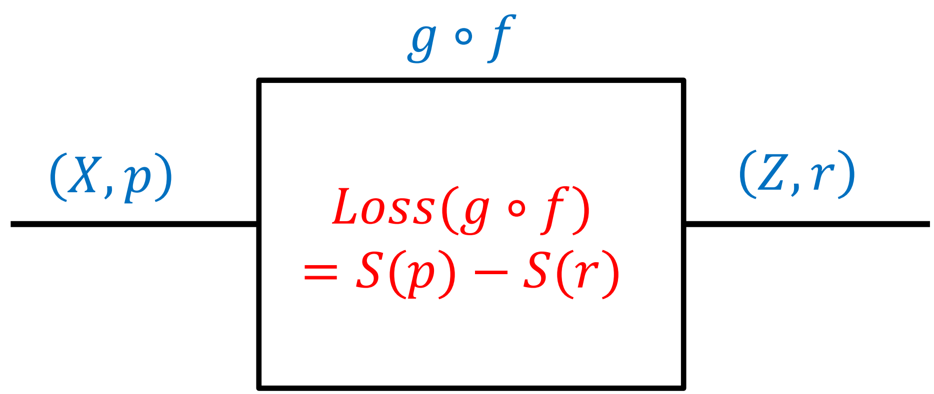


プロセスが「直列」に実行される場合

二つのプロセスが「直列」に実行されたとすると



それは、次のプロセスの実行と等しい。



プロセスが「直列」に実行される場合

この時、つぎの式が成り立つことがわかる。

$$
\begin{aligned}
Loss(g \circ f) &= S(p) - S(r) \\
&= \big(S(p) - S(q)\big) + \big(S(q) - S(r)\big) \\
&= Loss(f) + Loss(g)
\end{aligned}
$$

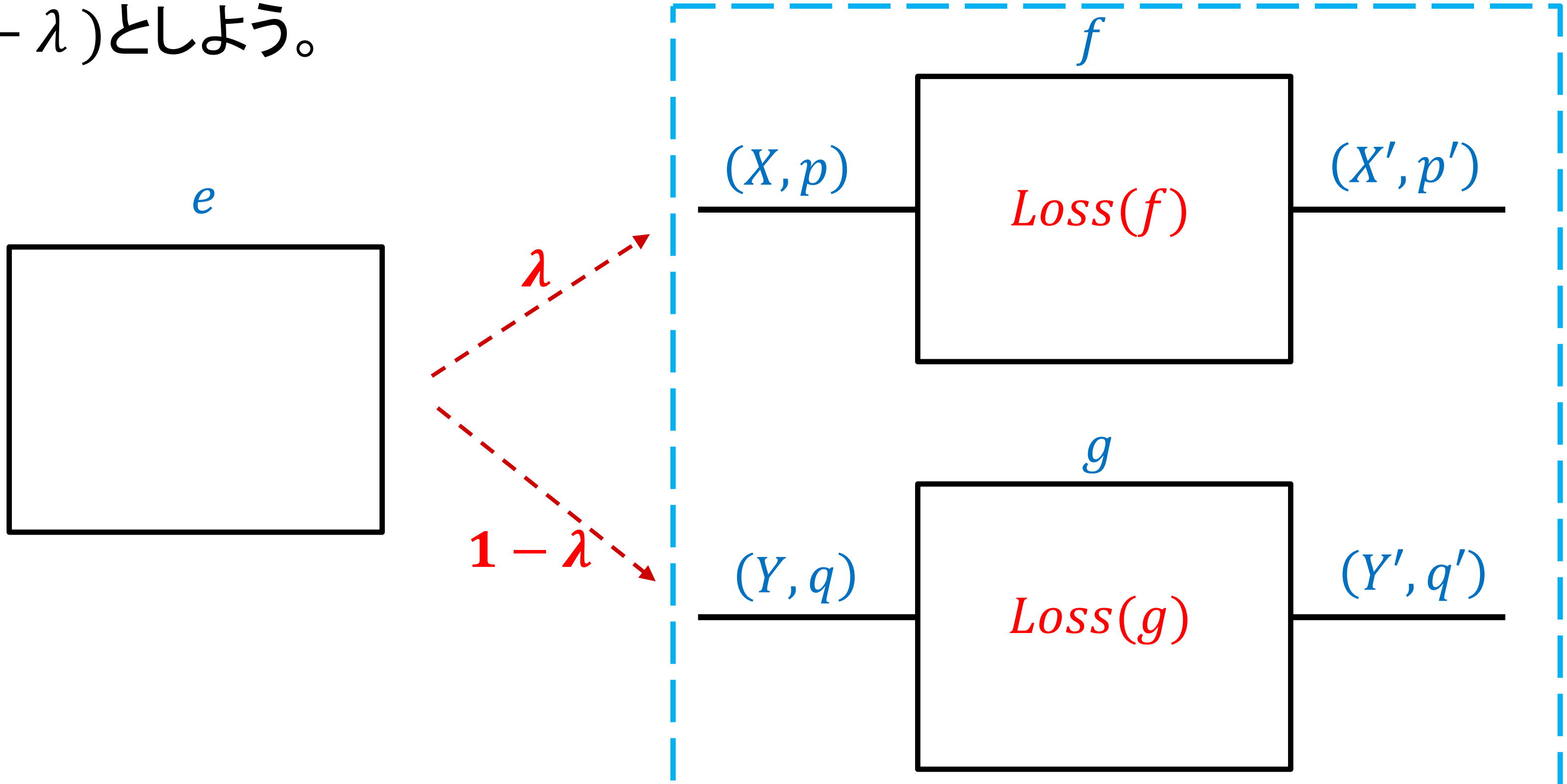
これは、次のように解釈できる。

二つの段階からなるプロセスを実行する時、全プロセスでの情報損失は、それぞれの段階での損失の和に等しい。

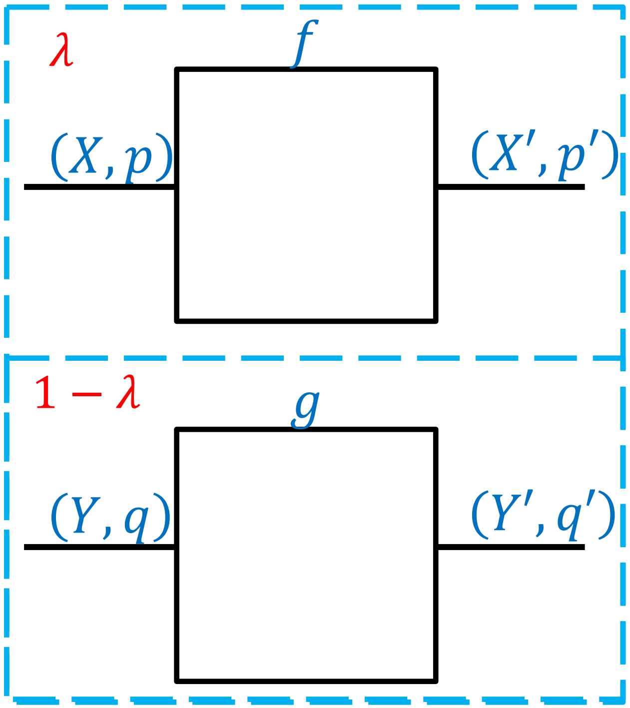
プロセスが「並列」に実行される場合

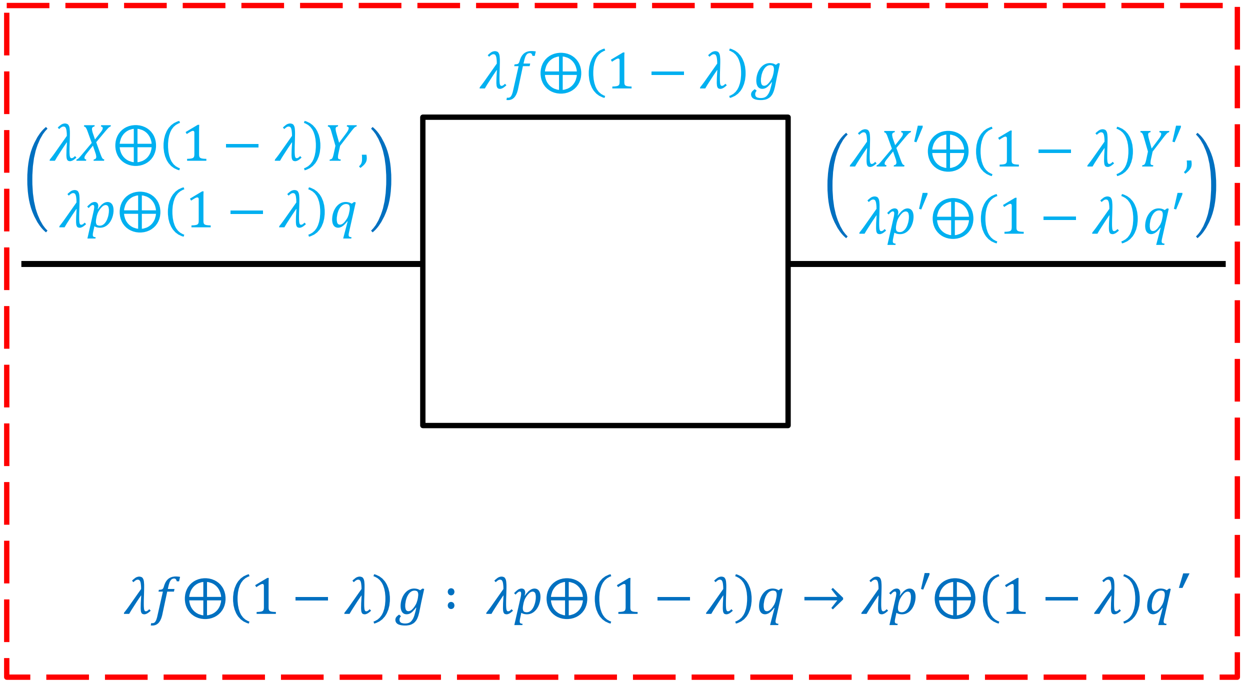
プロセスが二つに分岐して、二つのプロセスが「並行」に走る場合を考えてみよう。

プロセス f が選ばれる確率をλ、プロセス g が選ばれる確率を$(1-\lambda)$としよう。

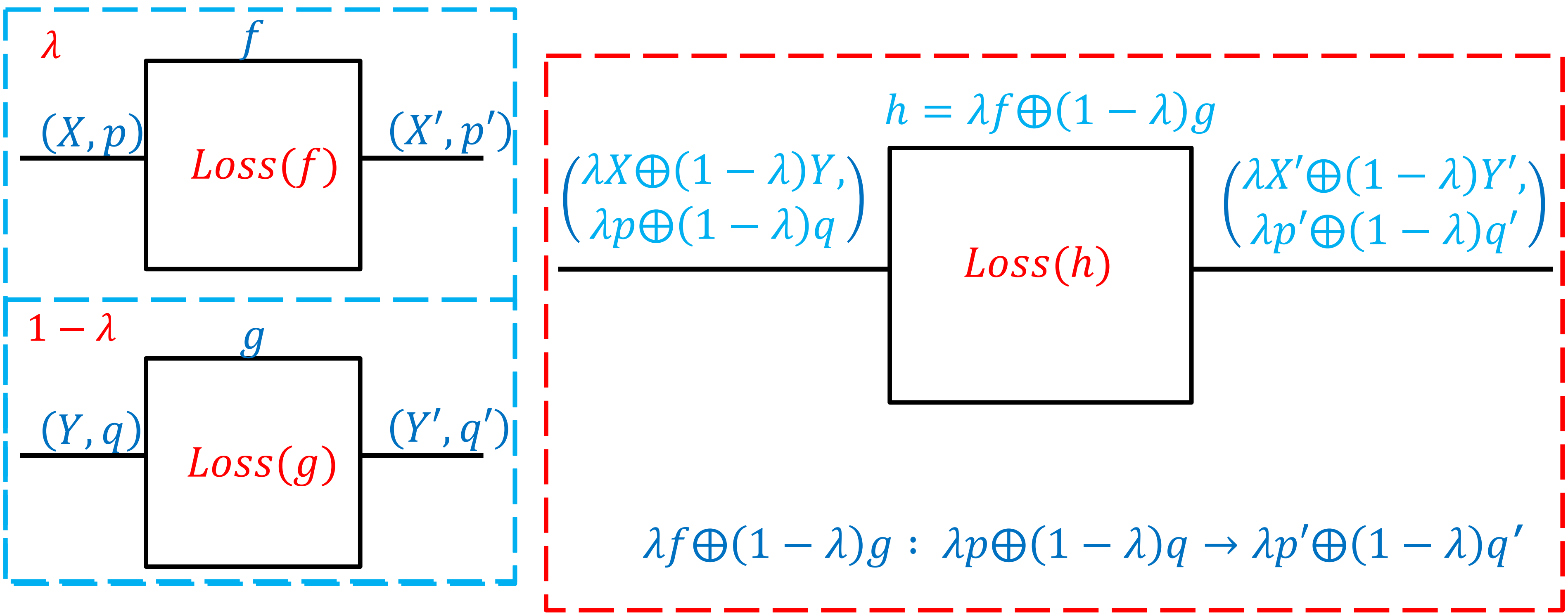


並列なプロセスを次のように表現する





並列なプロセスを次のように表現する



プロセスが「並列」に実行される場合の情報の損失 $Loss$ の計算

ここで、次の式が成り立つことを示せる。

$$Loss(\lambda f \oplus (1-\lambda)g) = \lambda Loss(f) + (1-\lambda)Loss(g)$$

これは、次のように解釈できる。

二つのプロセス f, g があった時、どちらを実行すべきかの選択を、表が出る確率がλのコインを投げて決めることにしよう。

この時、情報の損失は、f の損失のλ倍とg の損失の$(1-\lambda)$倍の和である。

「情報の損失」$Loss$の性質

$$Loss(g \circ f) = Loss(f) + Loss(g)$$

$$Loss(\lambda f \oplus (1-\lambda) g) = \lambda Loss(f) + (1-\lambda) Loss(g)$$

バエズが考えたこと　3

「情報の損失」の性質を公理化し、
シャノン・エントロピーの式を導く

「情報の損失」$Loss$の性質

$$Loss(g \circ f) = Loss(f) + Loss(g)$$

$$Loss(\lambda f \oplus (1-\lambda) g) = \lambda Loss(f) + (1-\lambda) Loss(g)$$

$Loss(f)$ は連続的

$Loss$が満たす公理を考える
公理 1

1) 測度を保存する関数fとgが合成可能ならいつでも、次の式が成り立つ。

$$F(f \circ g) = F(f) + F(g)$$

$Loss$が満たす公理を考える
公理 2

2. 確率測度p, q を持つ二つの有限集合と、実数$\lambda \in [0,1]$ があるとする。この時、二つの集合の互いに共通部分を持たない和集合上に確率測度$\lambda p \oplus (1-\lambda) q$が存在して、その値は、二つの測度を、それぞれ$\lambda$と$(1-\lambda)$で重み付けしたものである。

   同様に、射 $f: p \to p'$ と射 $g: q \to q'$ が与えられた時、あきらかに $\lambda p \oplus (1-\lambda) q$ から $\lambda p' \oplus (1-\lambda) q'$への射が存在する。この射を$\lambda f \oplus (1-\lambda) g$ と呼ぼう。

   この時、次の式が成り立つ。

$$F(\lambda f \oplus (1-\lambda) g) = \lambda F(f) + (1-\lambda) F(g)$$

$Loss$が満たす公理を考える
公理 3

3) 有限集合間の写像が「測度を保存する」ということは、それらの集合上の確率測度の与え方を変えることで、さまざまな異なる方法で考えることができる。こうした状況では、値$F(f)$は確率測度の与え方に依存する。

Fは、連続的である。

Fの連続性とは、実行するプロセスをほんの少し変えると、それが失う情報も、少しだけ変わるということである。

バエズが証明した定理

これら三つの条件をFが満たすとき、全ての測度を保存する関数$f:(X,p)\to(Y,q)$に対して、次の式を満たす定数$c\geq 0$が存在することを示すことができる。

$$F(f)=c\ Loss(f)=c(S(p)-S(q))$$

ここに、S はシャノン・エントロピーである。

別の言葉で言えば、この公理で定義される情報の損失$F(f)$は、シャノン・エントロピーの変化の定数倍$c\ Loss(f)$である。

この定理から、
シャノン・エントロピーの式を導くことができる

この定理から、シャノン・エントロピーの式を導くことができる。

$$S(p) = -\sum_{i \in X} p_i \log p_i$$

。

バエズの条件は、一意に シャノン・エントロピーを特徴づける

1. $F(f \circ g) = F(f) + F(g)$
2. $F(\lambda f \oplus (1-\lambda)g) = \lambda F(f) + (1-\lambda)F(g)$
3. Fは、連続的である

という三つの条件による「情報の損失」という特徴づけは、一意に、シャノン・エントロピーを特徴づける

バエズが考えたこと　4

Entropy as a Functor

Shannon

シャノンは、エントロピーHが、次の条件を満たすべきだと考えた。

1. H は、pi について連続的である。
2. もし全てのpiが等しく、pi ＝ 1/n であるとすれば、Hは、nについて単調増加の関数である。
（同じように見えるイベントが与えられ、より多くの可能なイベントが存在する時、より多くの選択、より多くの不確かさが存在することになる。）
3. もし選択が、連続的に行われる二つの選択に分解されるのなら、求めるHは、個々の選択のHの値の、重みづけられた和になるべきである。

Faddeev

Faddeevは、次の三つの公理から、シャノン・エントロピーの式が導かれることを示した。

1. $0 \le p \le 1$の時、$H(p, 1-p)$は連続的で、少なくとも一つの点で正である。
2. $H(p_1, p_2, \dots p_n)$は、$p_1, p_2, \dots p_n$について、対称的な関数である。
3. $n \ge 2$の時、$p_n = q_1 + q_2$として、
   $$H(p_1, p_2, \dots p_{n-1}; q_1, q_2) = H(p_1, p_2, \dots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$$

Leinster

Tom Leinster は、Faddeevの条件から出発して、さらに、次のことを証明した。

シャノン・エントロピーをHとし、$(I : \Delta_n \to \mathbb{R})_{n\geq 1}$を、確率分布を実数に写す関数列とする。この時、次の二つの条件は等しい。

i. 関数 Iは連続で、Chain ルールを満たす

$$I\big(\boldsymbol{w} \circ (\boldsymbol{p}^1, \dots, \boldsymbol{p}^n)\big) = I(\boldsymbol{w}) + \sum_{i=1}^{n} w_i\, I(p^i)$$

ii. ある $c \in \mathbb{R}$ について、$I = cH$

Baez

1. $F(f \circ g) = F(f) + F(g)$
2. $F(\lambda f \oplus (1-\lambda) g) = \lambda F(f) + (1-\lambda) F(g)$
3. Fは、連続的である

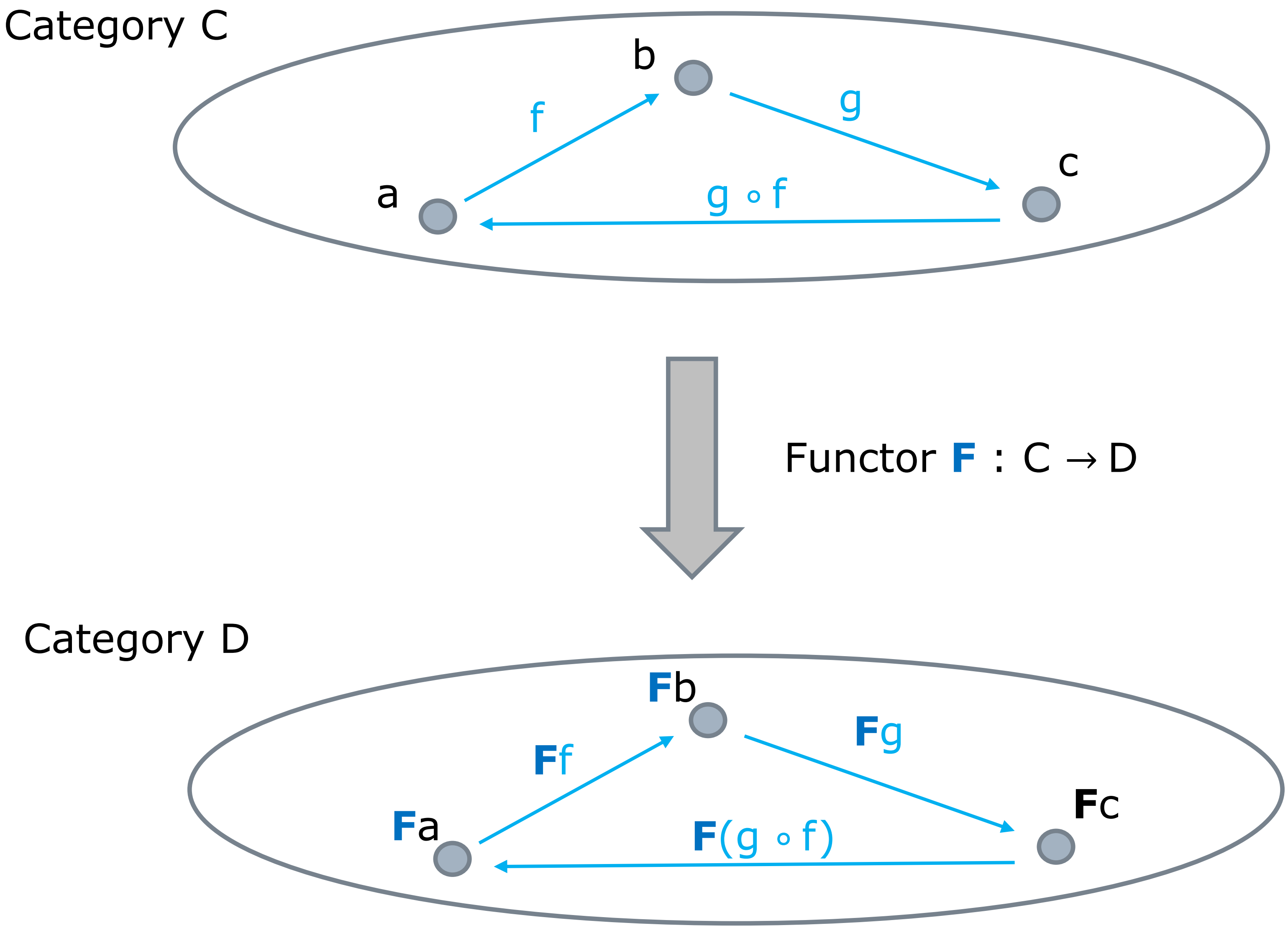
という三つの条件による「情報の損失」という特徴づけは、
一意に、シャノン・エントロピーを特徴づける

Baez
Entropy as a Functor

1. $F(f \circ g) = F(f) + F(g)$
2. $F(\lambda f \oplus (1-\lambda)g) = \lambda F(f) + (1-\lambda)F(g)$
3. Fは、連続的である

1. Fは、Functor である
2. Fは、convex linearなFunctorである
3. Fは、連続的なFunctorである

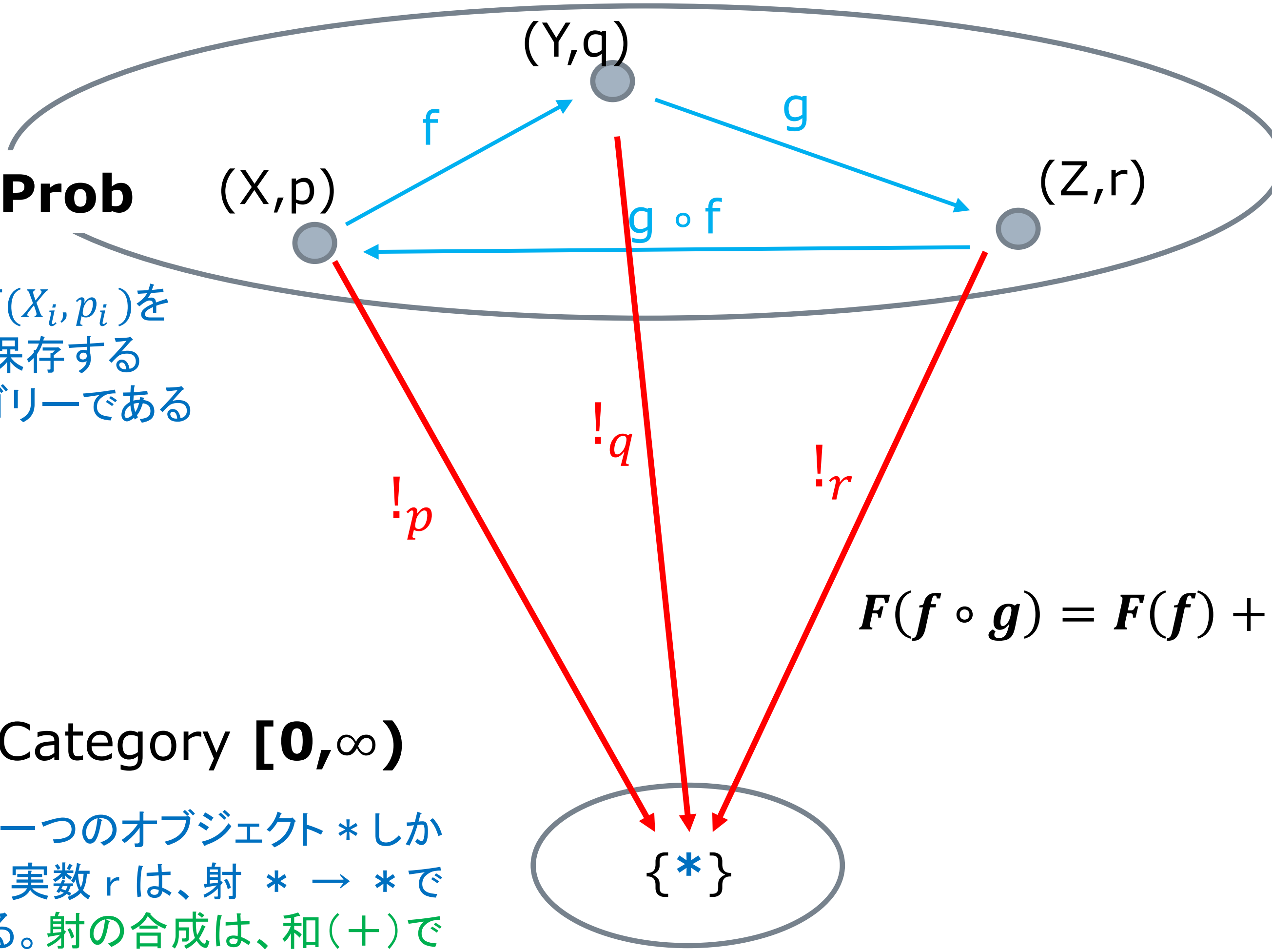
CategoryとFunctor



Functor **F** : **FinProb** → **[0,∞)**

Category **FinProb**

FinProbは、確率分布(X_i, p_i)をオブジェクト、測度を保存する関数を射とするカテゴリーである

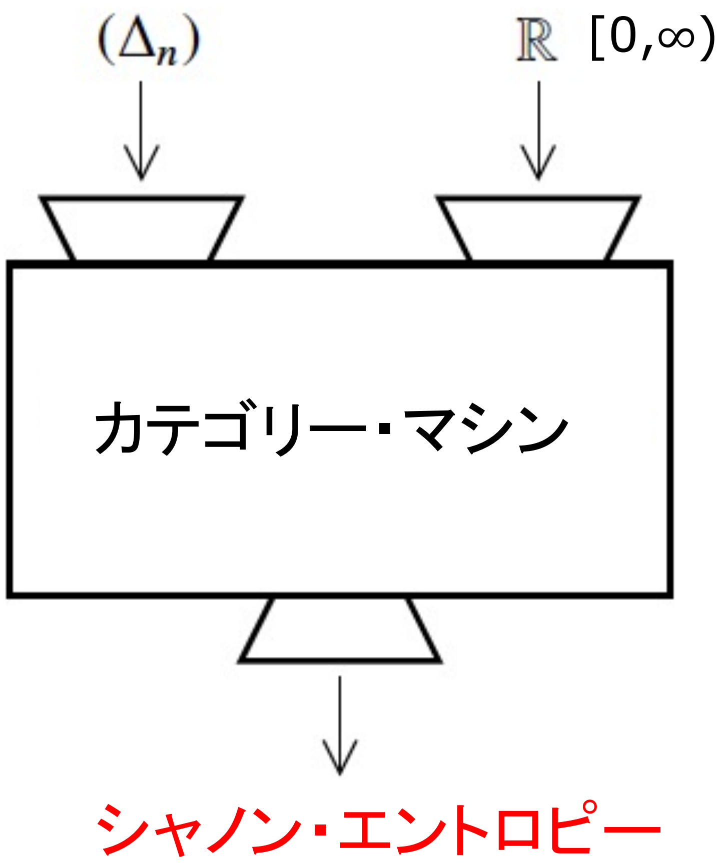


$$\boldsymbol{F(f \circ g) = F(f) + F(g)}$$

Category **[0,∞)**

[0,∞) は、一つのオブジェクト＊しか持たない。実数 r は、射 ＊ → ＊ で表現される。射の合成は、和（＋）で定義される

カテゴリー論の魅力と威力



入力に、シンプレックス(Δ_n)と
実数軸が与えられれば、
出力にシャノン・エントロピー
の概念を返す、一般的な能力
を持つカテゴリー論的マシン
が存在する！