エントロピー論とカテゴリー論

エントロピー論とカテゴリー論
Agenda

第一部
新しいエントロピー論の登場

第二部
カテゴリー論的アプローチの基礎

第三部
相対エントロピーのカテゴリー論的解釈

Appendix A
エントロピー概念の拡大

第一部
新しいエントロピー論の登場

1. 代表的なエントロピーの特徴づけ
2. 「情報の損失」としてのエントロピー
3. 測度を保存する関数について
4. 「情報の損失」の性質を考える
5. Entropy as a Functor

第二部
カテゴリー論的アプローチの基礎

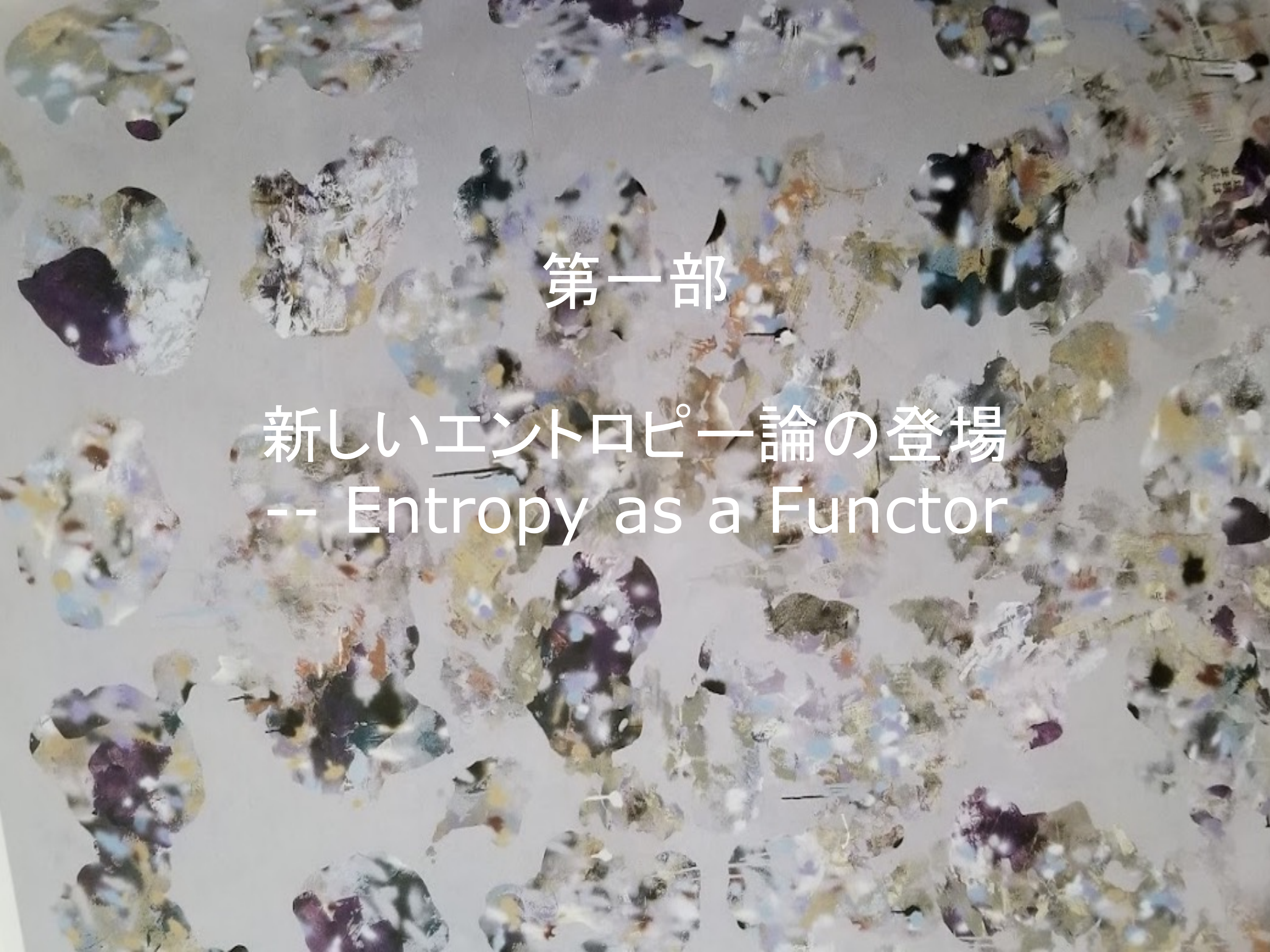
1. 1 への射 – カテゴリーとしての実数
2. シャノン･エントロピーのカテゴリー論的解釈の設定
3. stochastic map
4. measure preserving function
5. 1 からの射 $1 \rightarrow X$
6. 確率分布の表現
7. 可換な図形
8. 可換な図形が表現するもの
9. 相対エントロピーのカテゴリー論的解釈の設定

第三部
相対エントロピーのカテゴリー論的解釈

1. 相対エントロピー
2. stochastic map
3. stochastic mapとFinProb
4. Finstatと「仮説」
5. Functorとしての相対エントロピー

Appendix A
エントロピー概念の拡大

1. Tsallis エントロピー
2. カテゴリーとして実数を捉える
3. Tsallis エントロピー = q-$\boldsymbol{logarithmic}$エントロピー
4. RényiエントロピーとHill数



第一部

新しいエントロピー論の登場
-- Entropy as a Functor

第一部
新しいエントロピー論の登場

1. 代表的なエントロピーの特徴づけ
2. 「情報の損失」としてのエントロピー
3. 測度を保存する関数について
4. 「情報の損失」の性質を考える
5. Entropy as a Functor

代表的なエントロピーの特徴づけ

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. $0 \leq p \leq 1$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、
$$H(p_1, p_2, \ldots p_{n-1}; q_1, q_2) = H(p_1, p_2, \ldots, p_n) + p_n H(\frac{q_1}{p_n}, \frac{q_2}{p_n})$$

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

「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を紹介し、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 $f\colon (X,p) \to (Y,q)$に対して、次が言える。

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

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。

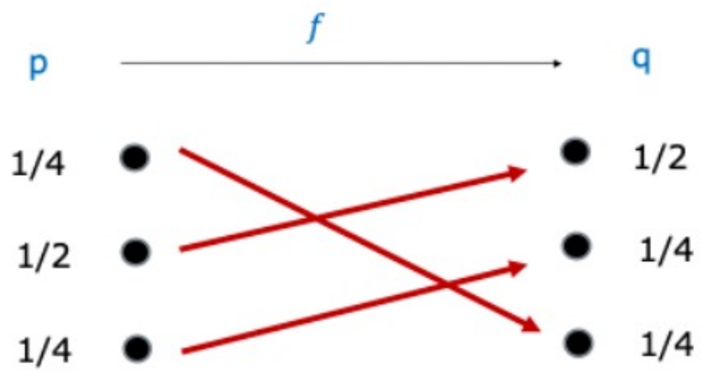
$$f:p\to q$$
$$F(f)=F(p)-F(q)$$

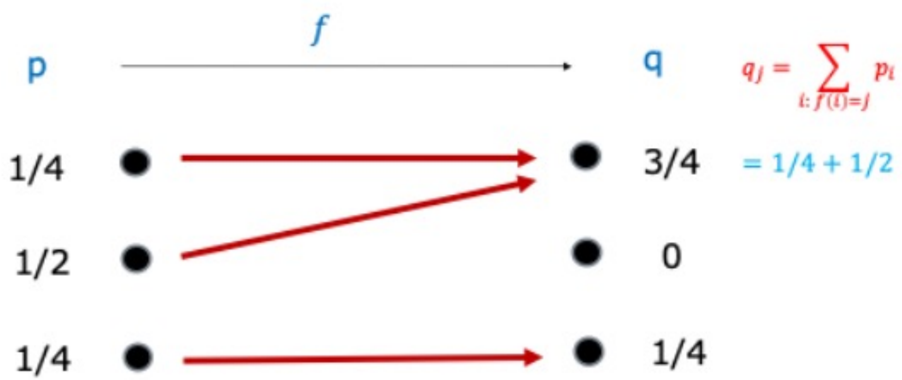
それを、我々は「情報の損失」と考える。

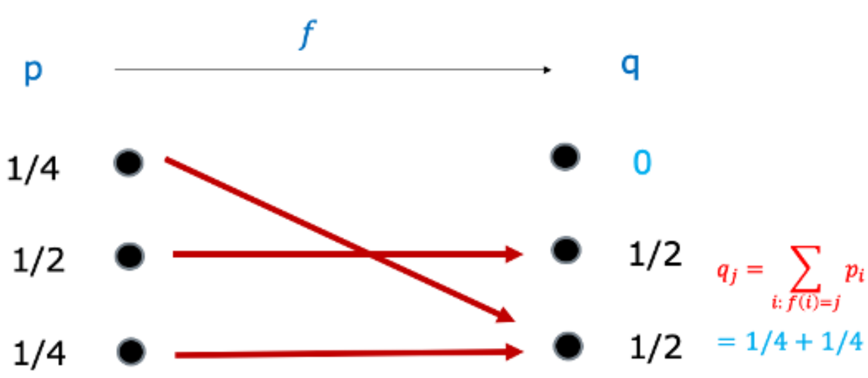
ここでは、シャノンのエントロピーを「情報の損失」という言葉で特徴づけようと思う。

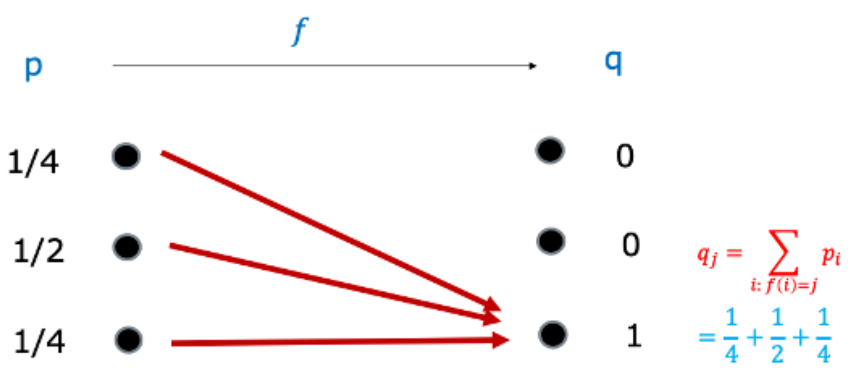
測度を保存する関数

測度を保存する関数 f の例

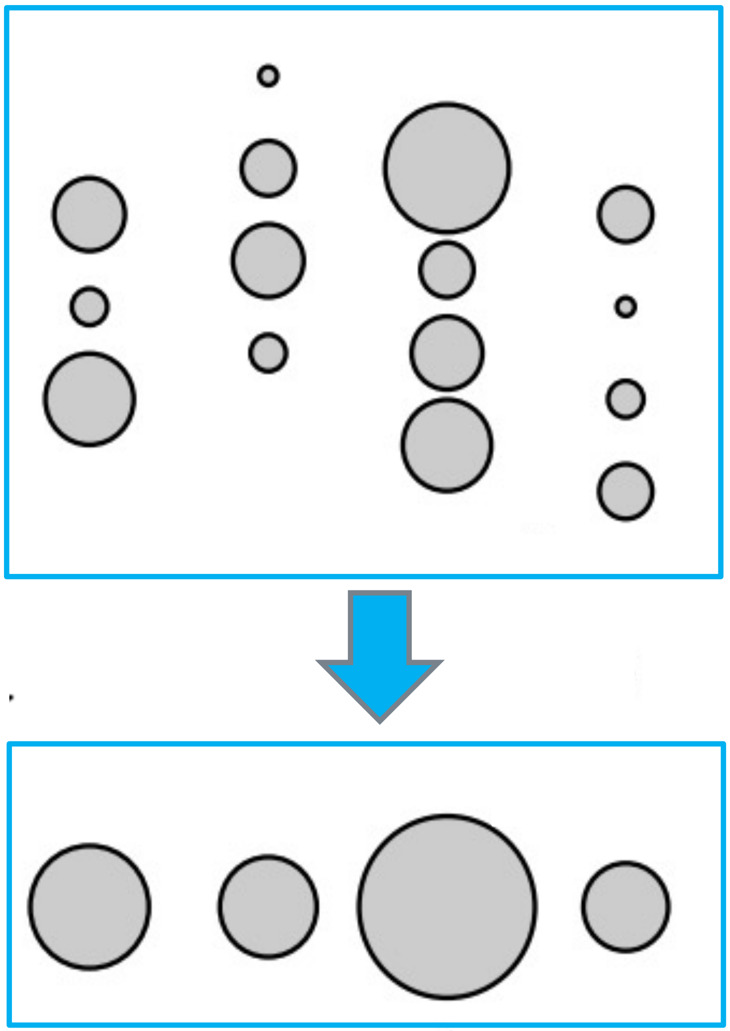








測度を保存する関数 f の例
水滴モデル

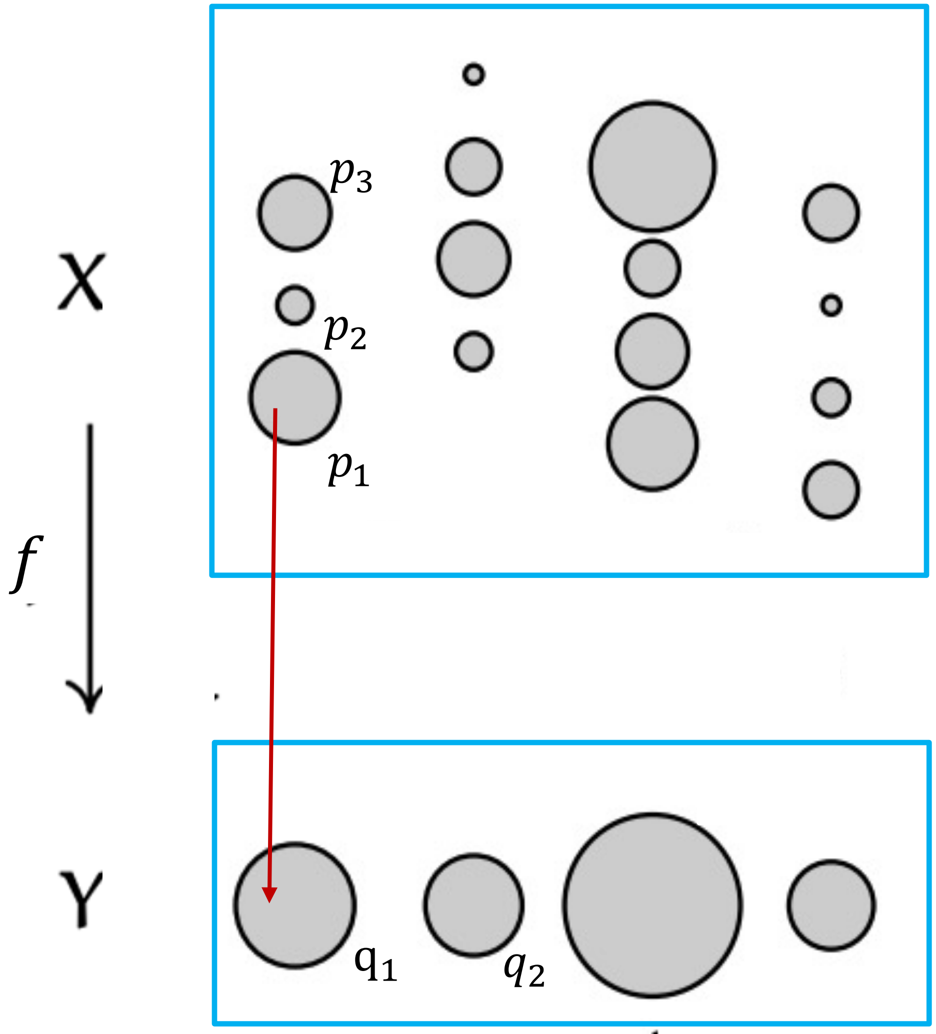


Xは水滴の集まりである。水滴は落下して 豪流して形を変え、水滴の集まりYになったとする。

Xの水滴の体積は、全部で1だとする。この時、Yの水滴の体積も1である。

（要素の体積を全部足して1になるというう想定は、は、要素は確率を表し。X,Yは確率分布であることを表現している。）

fは、Xの要素p_iにYの要素q_jをひとつ対応させる



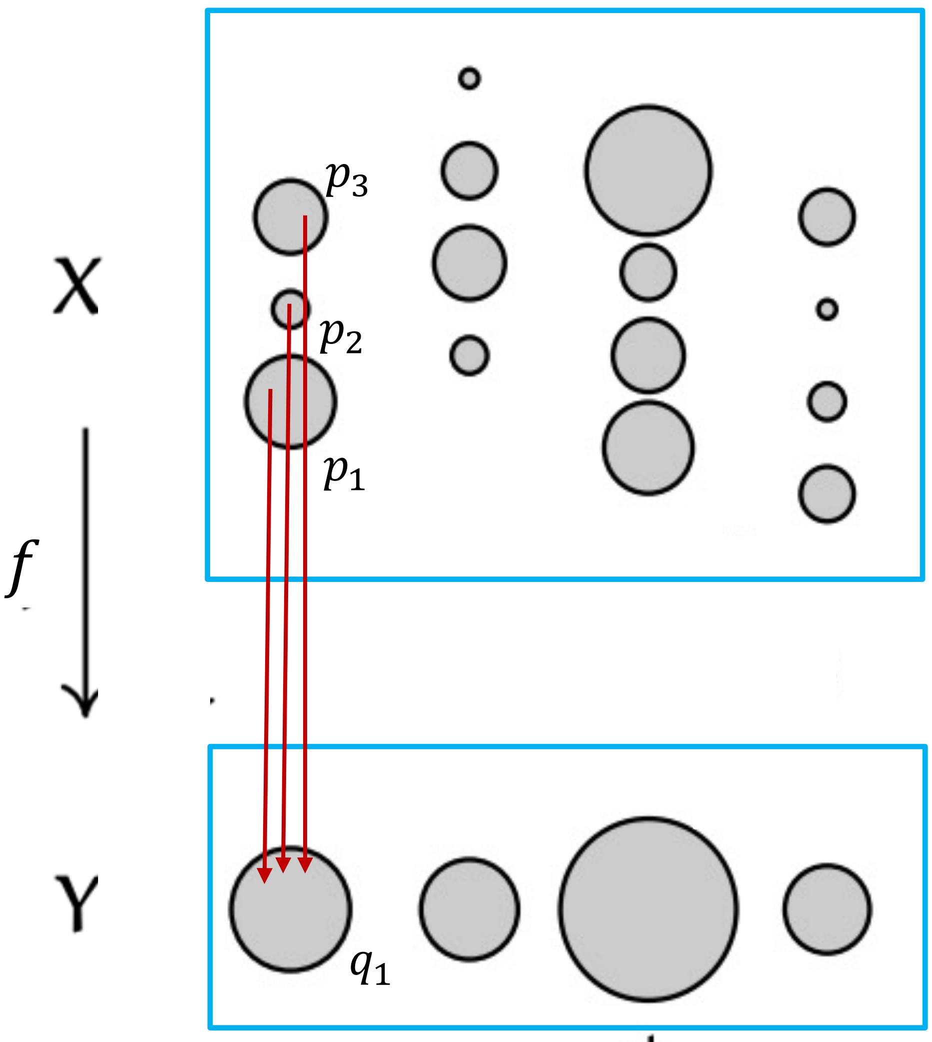
fを、確率分布(X, p)から確率分布(Y, q)への関数とする。fは、Xの要素p_iに、Yの要素q_jをひとつ対応させる。

$$x_i \to q_j$$

これを次のように表す。

$$f(i) = j$$

Xの水滴は合流する



左の図でXで縦に一直線に並んでいる水滴は、下のYの水滴に合流するものとする。

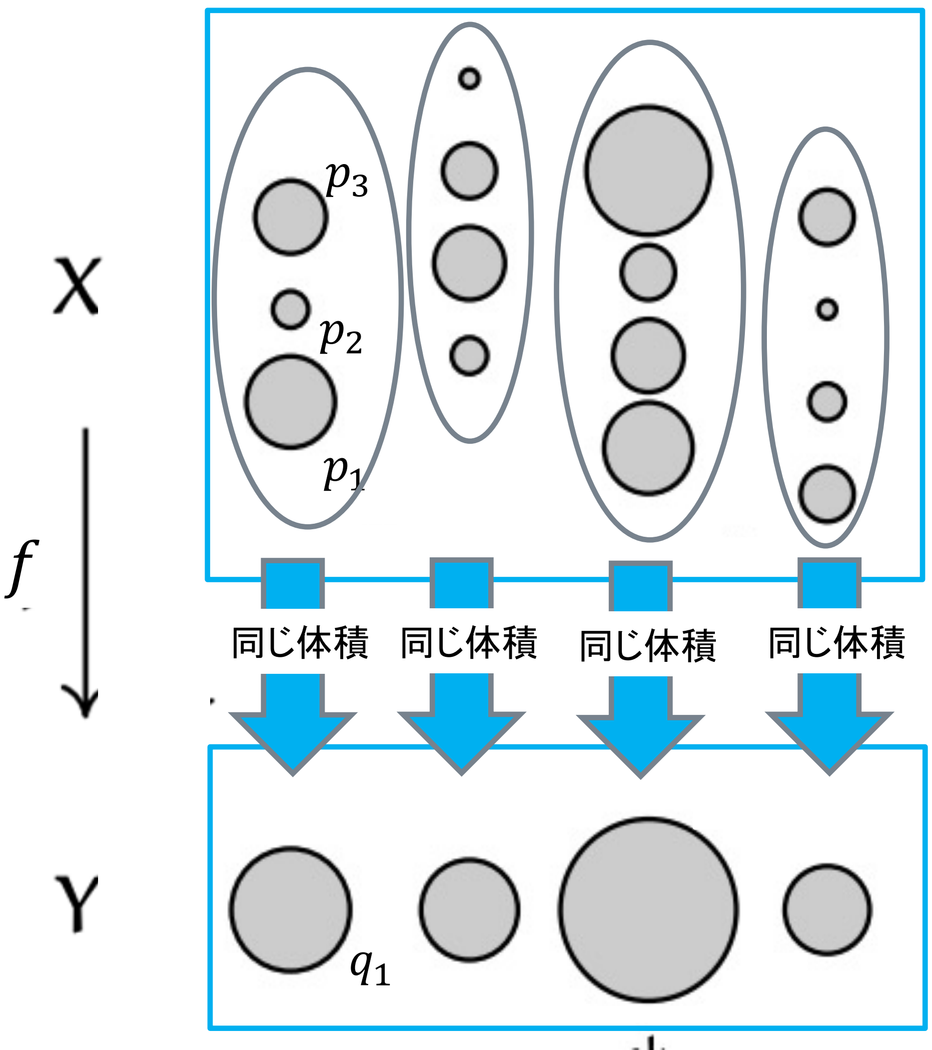
左の図でのx_1, x_2, x_3の水滴は、水滴y_1に合流している。

式で表せば、

$$f(1) = 1$$
$$f(2) = 1$$
$$f(3) = 1$$

と表現できる。

合流によって体積は変わらない



合流後の水滴の体積は、合流した水滴の体積を足したものである。

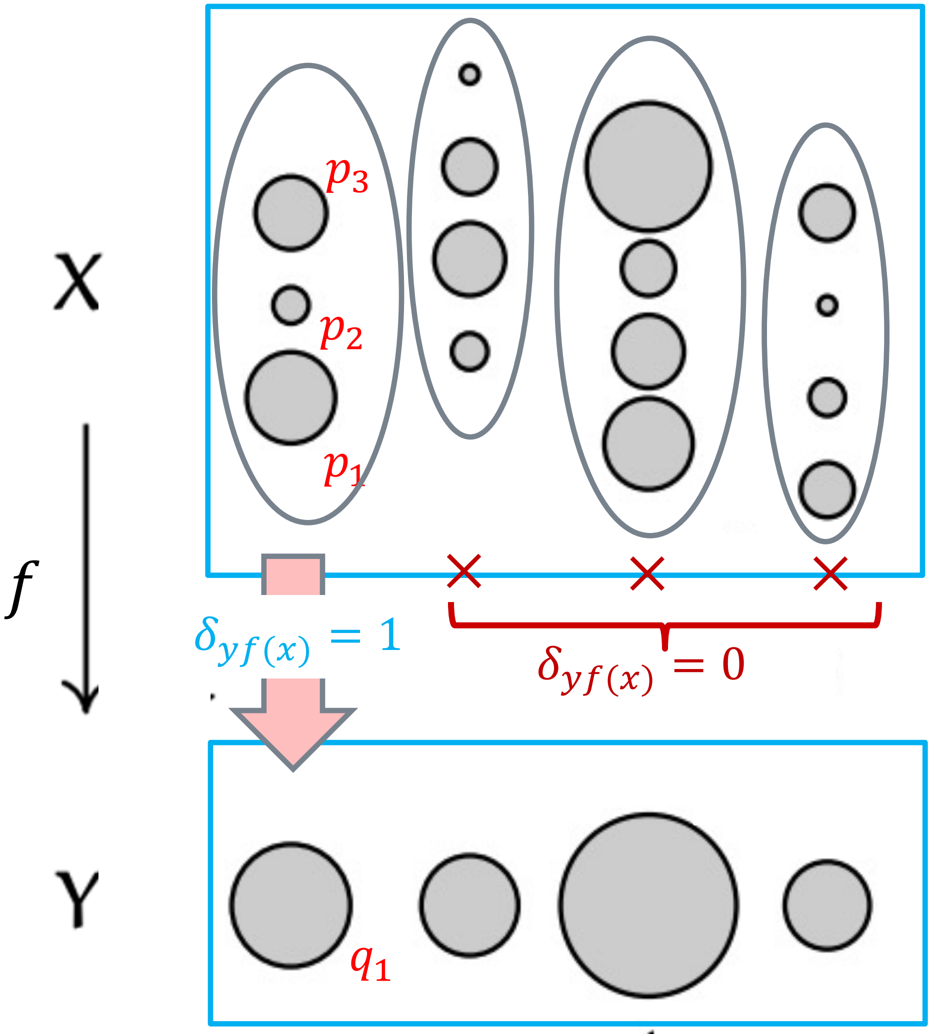
全ての縦の列について、こうした関係が成り立つ。

変数の名前で水滴の体積を表すことにすれば、左の例だと

$$p_1 + p_2 + p_3 = q_1$$

となる。

測度を保存する関数のイメージ



合流前後で体積は変わないことを式で表す

$f(i) = j$である全てのp_iを足したものがq_jとなるので、次の式がなりたつ。

$$q_j = \sum_{x \in X} \delta_{yf(x)} p_x$$

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

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。

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

「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。(前回セミナーの第二部)
- 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。（前回セミナーの第三部）

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。

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

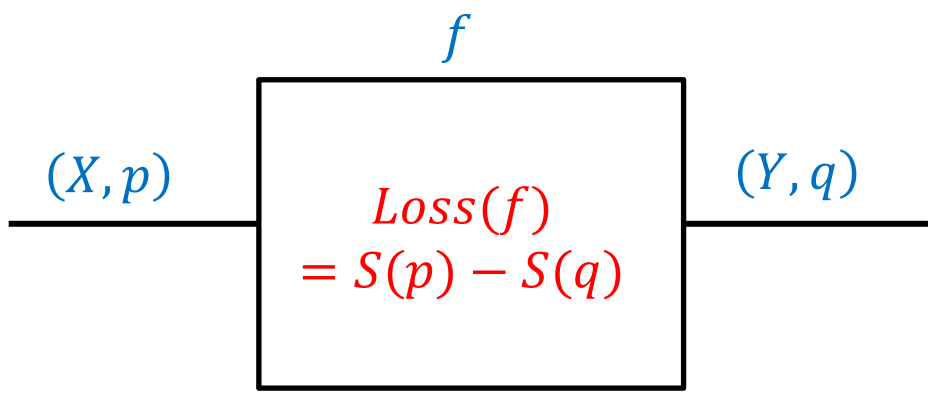
$S(p), S(q)$ は、確率分布 p, q のシャノン・エントロピー

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。

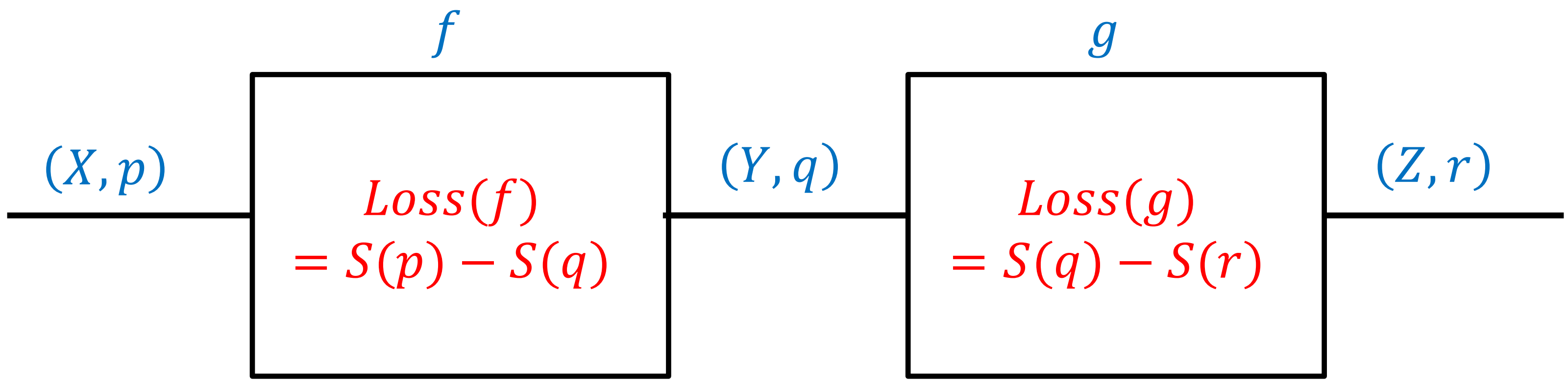
「情報の損失」をプロセスとして見る

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。

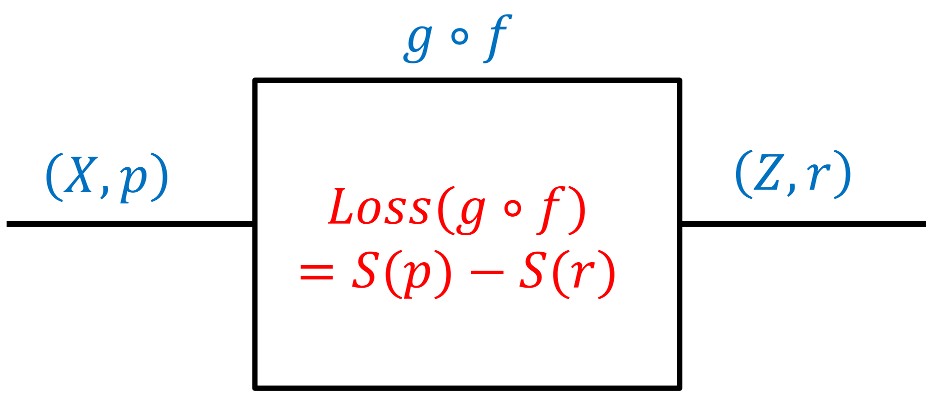


プロセスが「直列」に実行される場合

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。

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

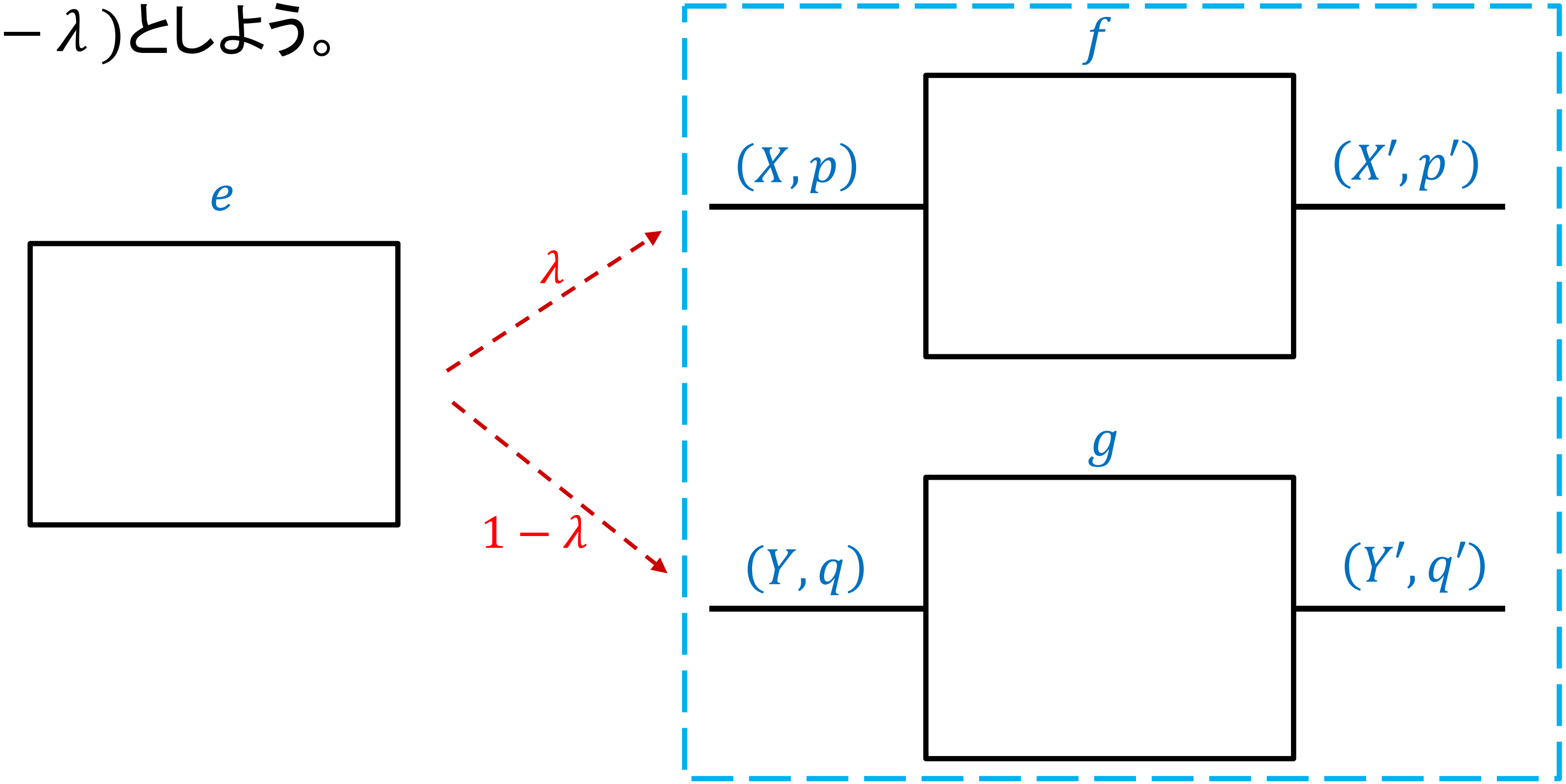
これは、次のように解釈できる。

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。



確率分布$(X,p),(Y,q)$から
確率の集合 $\lambda p \oplus (1-\lambda)q$ を構成する

互いに共通部分を持たない集合X,Y 上に、それぞれ確率分布 $(X,p),(Y,q)$が与えられているとする。

実数 $\lambda \in [0,1]$に対して、確率の集合 r を次のように定義する。

$k \in \mathrm{X}$の時、$r_k = \lambda p_k$
$k \in Y$の時、$r_k = (1-\lambda)q_k$

この r を次のように表す。

$$r = \lambda p \oplus (1-\lambda)q$$

確率分布$(X,p),(Y,q)$から
確率分布 $(X+Y,\lambda p\oplus(1-\lambda)q)$を構成する

集合$X\oplus Y$上で定義された確率の集合 $r=\lambda p\oplus(1-\lambda)q$ に対して、

$$\sum_k r_k=\sum_i \lambda p_i+\sum_j (1-\lambda)q_j=\lambda\sum_i p_i+(1-\lambda)\sum_j q_j$$
$$=\lambda\times 1+(1-\lambda)\times 1=1$$

よって、

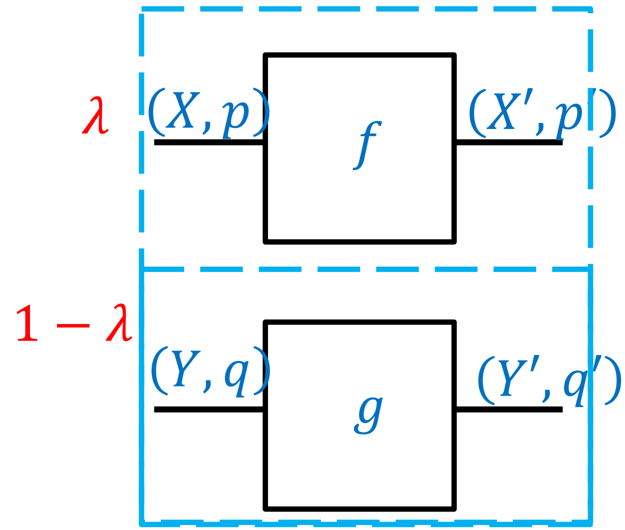
$$(X\oplus Y,r)=(X+Y,\lambda p\oplus(1-\lambda)q)$$

は、集合$X+Y$上の確率分布である。

（まぎれがなければ、集合$X\oplus Y$を、$X+Y$で表しても構わない。）

$f\colon (X,p) \to (X',p')$

$g\colon (Y,q) \to (Y',q')$



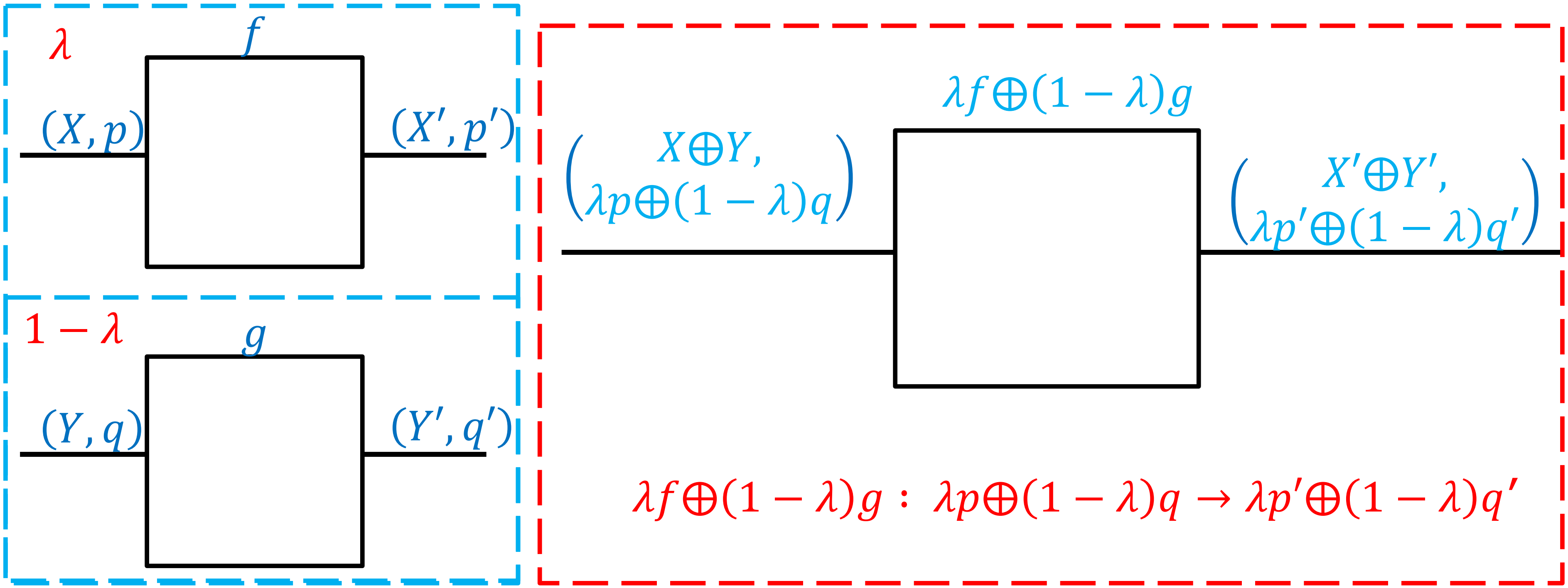
$f\colon (X,p) \to (X',p')$, $g\colon (Y,q) \to (Y',q')$ とすると

$$\lambda \left\{ \begin{array}{c} (X,p) \\ \downarrow f \\ (X',p') \end{array} \right\} \oplus (1-\lambda) \left\{ \begin{array}{c} (Y,q) \\ \downarrow g \\ (Y',q') \end{array} \right\} = \begin{array}{c} (X+Y, \lambda p \oplus (1-\lambda)q) \\ \downarrow \\ (X'+Y', \lambda p' \oplus (1-\lambda)q') \end{array}$$

$\lambda f \oplus (1-\lambda) g$:

$$(X+Y, \lambda p \oplus (1-\lambda)q) \to (X'+Y', \lambda p' \oplus (1-\lambda)q')$$

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。

Entropy as a Functor

「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
もっと一般なFで　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 もっと一般なFで　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 もっと一般なFで　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。

別の言葉で言えば、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

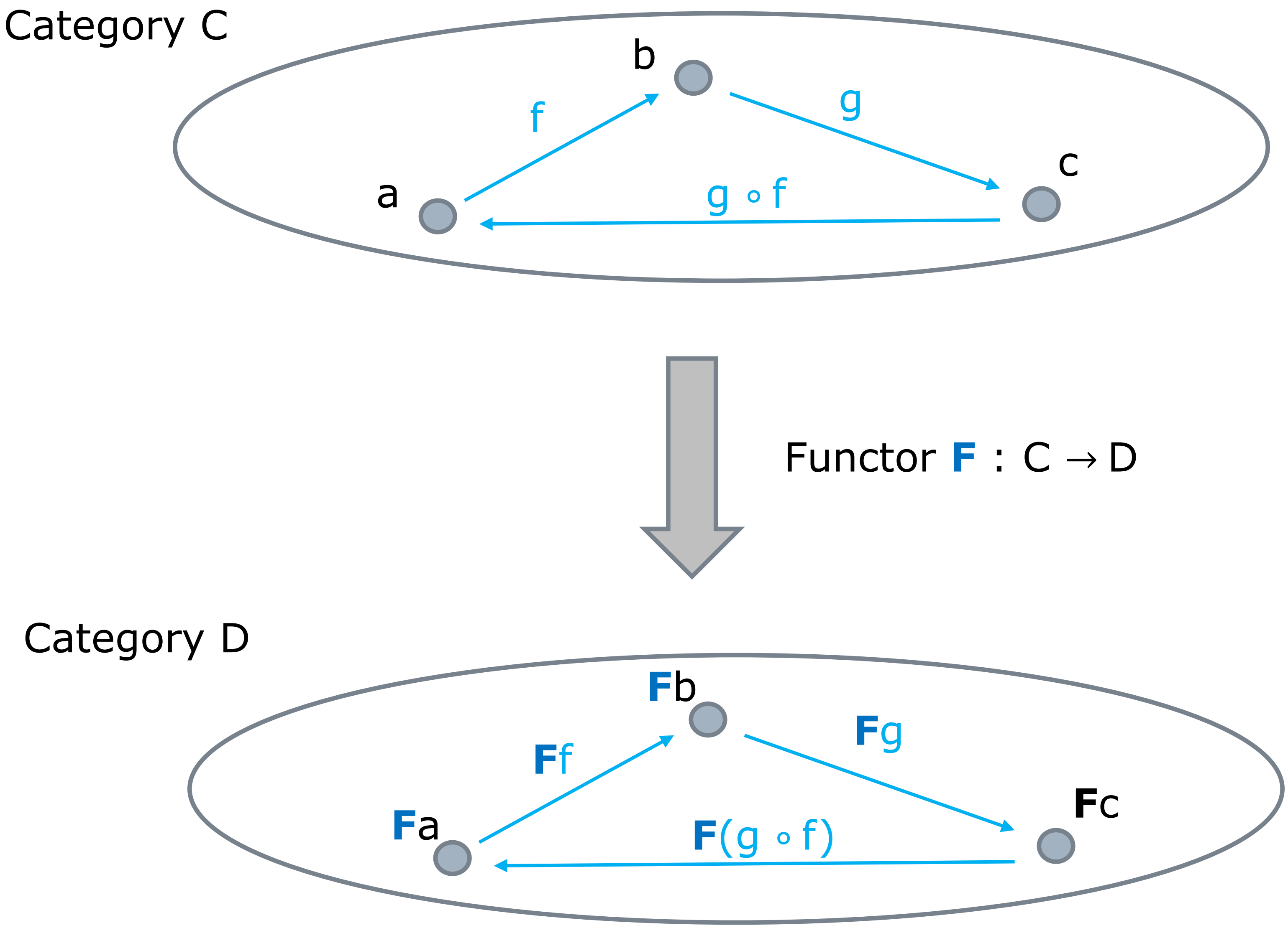
という三つの条件による「情報の損失」という特徴づけは、一意に、シャノン・エントロピーを特徴づける

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

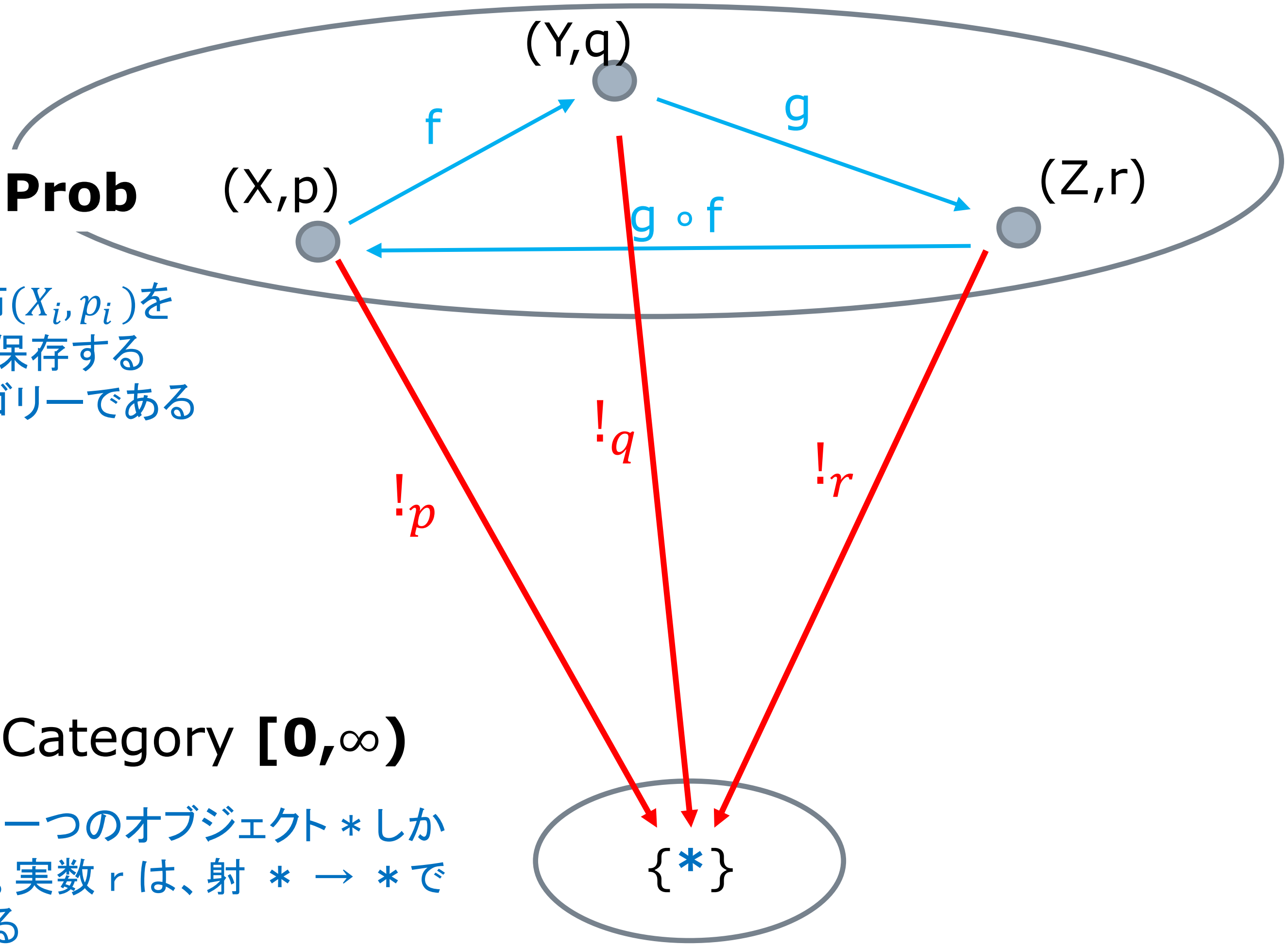
1. Fは、Functor である
2. Fは、convex linearなFunctorである
3. Fは、連続的なFunctorである

CategoryとFunctor



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

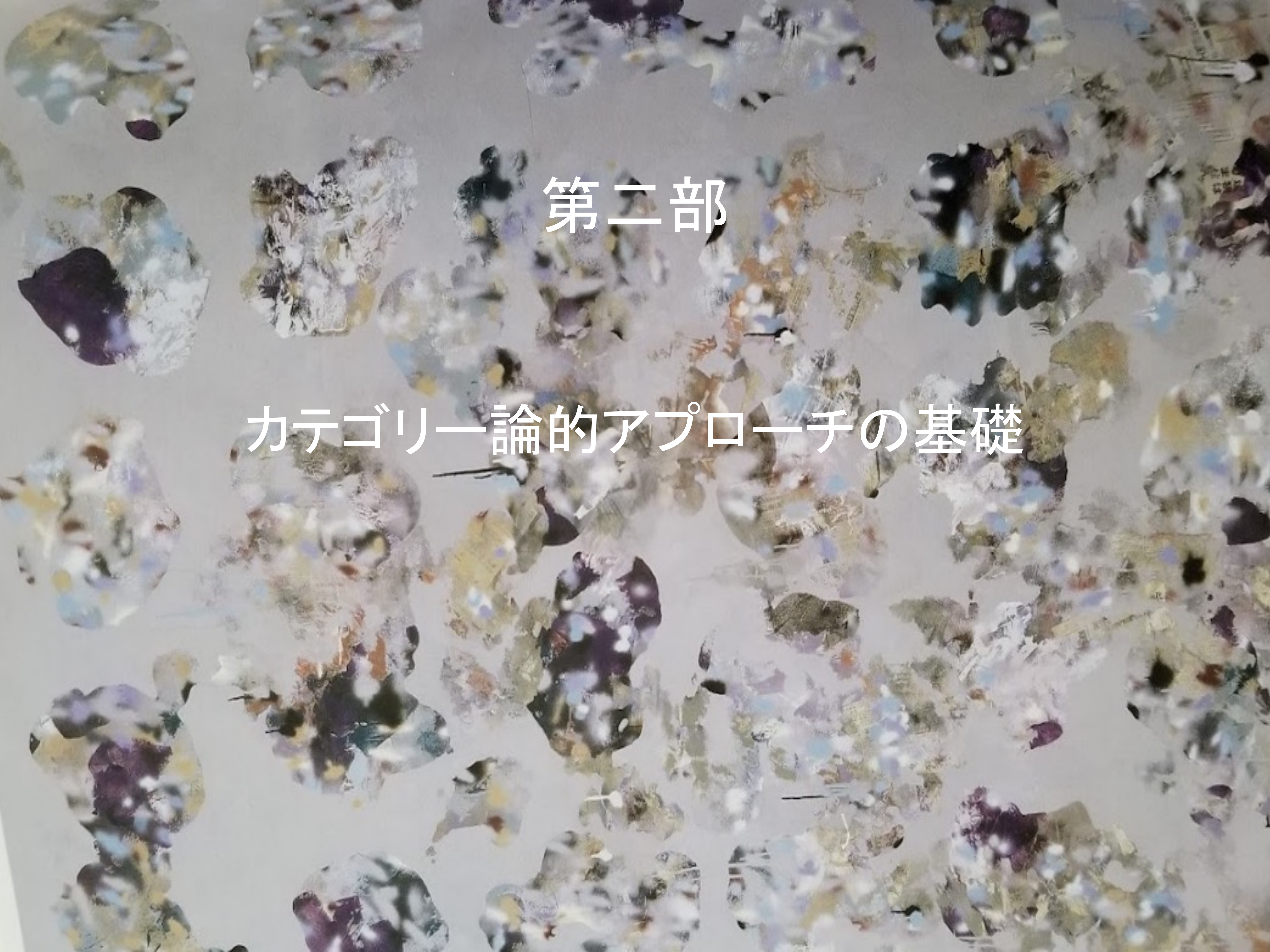
Category **FinProb**



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

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



第二部

カテゴリー論的アプローチの基礎

第二部
カテゴリー論的アプローチの基礎

1. 1 への射 – カテゴリーとしての実数
2. シャノン･エントロピーのカテゴリー論的解釈の設定
3. stochastic map
4. measure preserving function
5. 1 からの射 $1 \to X$
6. 確率分布の表現
7. 可換な図形
8. 可換な図形が表現するもの
9. 相対エントロピーのカテゴリー論的解釈の設定

1 への射 – カテゴリーとしての実数

Shannonエントロピーが Functorであるという議論を振り返る

ShannonエントロピーがFunctorであるという議論をしてきたのだが、そこでの重要な論点は、エントロピーが取る非負の実数 $[0,\infty)$という値の範囲そのものが、カテゴリーとみなすことができるということである。ただ、実数をカテゴリーとして捉えることができるということは、自明ではない。

確率分布のカテゴリーから $[0,\infty)$というカテゴリーへの、ある性質を満たすFunctor Fは、シャノン・エントロピーを一意に決定するという発見が重要なのだが、ここでは、「実数というカテゴリー」の理解にフォーカスして、その逆の、シャノン・エントロピーの差で表現される「情報の損失」は、Functor とみなせることを、もう少し詳しくみておこう。

一つのオブジェクトしかもたない確率分布 $(\{1\},\{1\})$

一つのオブジェクトしかもたない確率分布$(\{*\}, p)$を考えてみよう。これは、確率変数が一つの確率分布$(\{1\}, p)$と表してもいい。

一つしか確率変数がないのだから、pは1になる。一つのオブジェクトしかもたない確率分布は、$(\{1\}, \{1\})$と表すことができる。

確率分布(X, p)から一つのオブジェクトしかもたない確率分布$(\{1\}, \{1\})$への測度を保存する関数を考え、それを $!_p$ と表そう。

確率分布(X, p)から確率分布$(\{1\}, \{1\})$への関数$!_p$

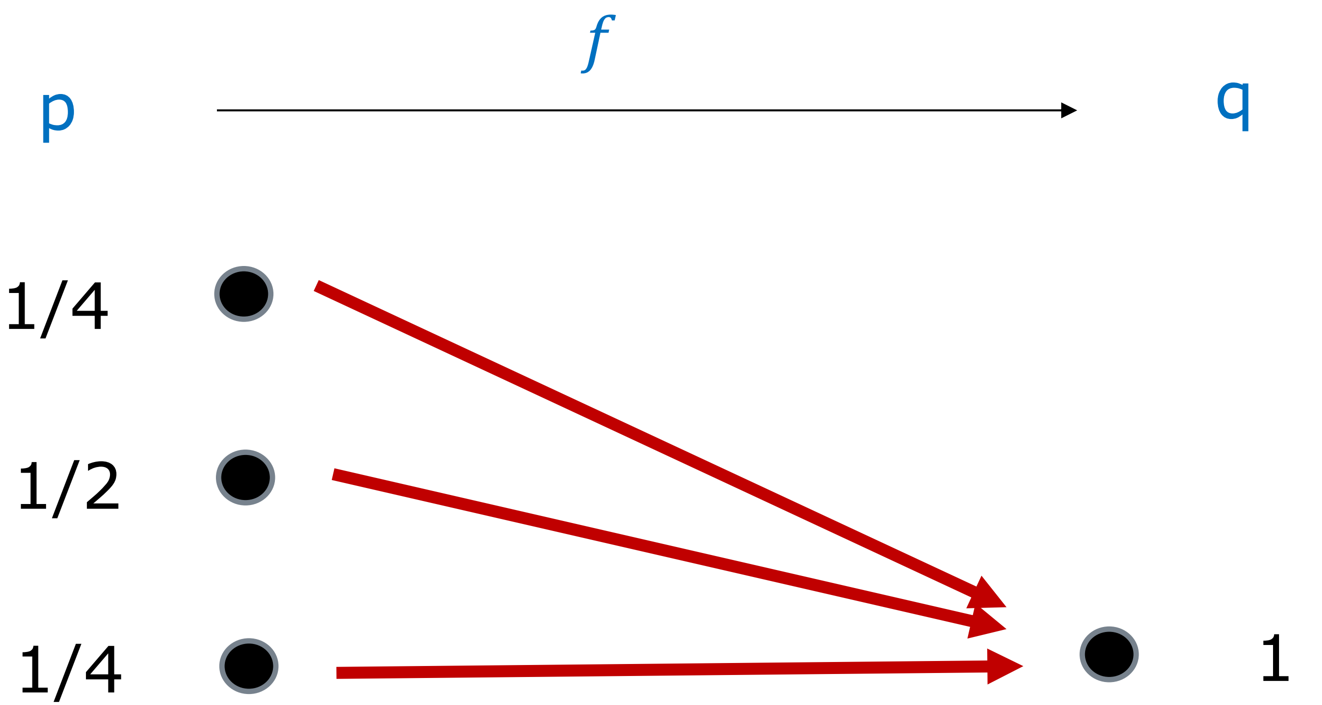
例えば、フェアなコインの確率分布は、({1, 2}, {1/2, 1/2})と表せるのだが、それから ({1}, {1})への関数を考えてみる。

この場合、関数 $!_p$: ({1, 2}, {1/2, 1/2}) → ({1}, {1})は、コインに裏表 {1, 2} があることも、それぞれが等しい確率{1/2, 1/2}を持つことも、すべて忘れることになる。

ある確率分布(X, p)からエントロピーHを得る時、我々は同じことをしている。元の分布が，何個の確率変数を持ち、それぞれがどういう確率を持っていたのかは、捨象され忘れられているから。

関数 $!_p$ の例

次の測度を保存する関数 f は、確率分布(X, p)から$(\{1\}, \{1\})$への関数 $!_p$ の例である。



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

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

測度を保存する関数fと、$!_p$ の関係

測度を保存する任意の関数 f について、次の図式は可換である。

$$\begin{tikzcd}
(X,p) \arrow[rr, "f"] \arrow[dr, "!_p"'] & & (Y,q) \arrow[dl, "!_q"] \\
& (\{1\},\{1\}) &
\end{tikzcd}$$

すなわち、$!_q \circ f = !_p$ 。

I_p, I_qも測度を保存する関数である。

cを正の定数、$H(p)$をシャノン・エントロピーとした時、

$$F\left(!_p\right) = cH(p)$$

と定義する。

Fは、測度を保存する関数 $!_p$ を、シャノン・エントロピーの定数倍の実数$[0, \infty)$に写す関数である。

この時、Fは、次の定義のもとで、カテゴリー$FinProb$から、カテゴリー $[0, \infty)$へのFunctor である。

また、$FinProb$の任意の射 $f: p \to q$に対して、次の式が成り立つ。

$$F(f) = c(H(p) - H(q))$$

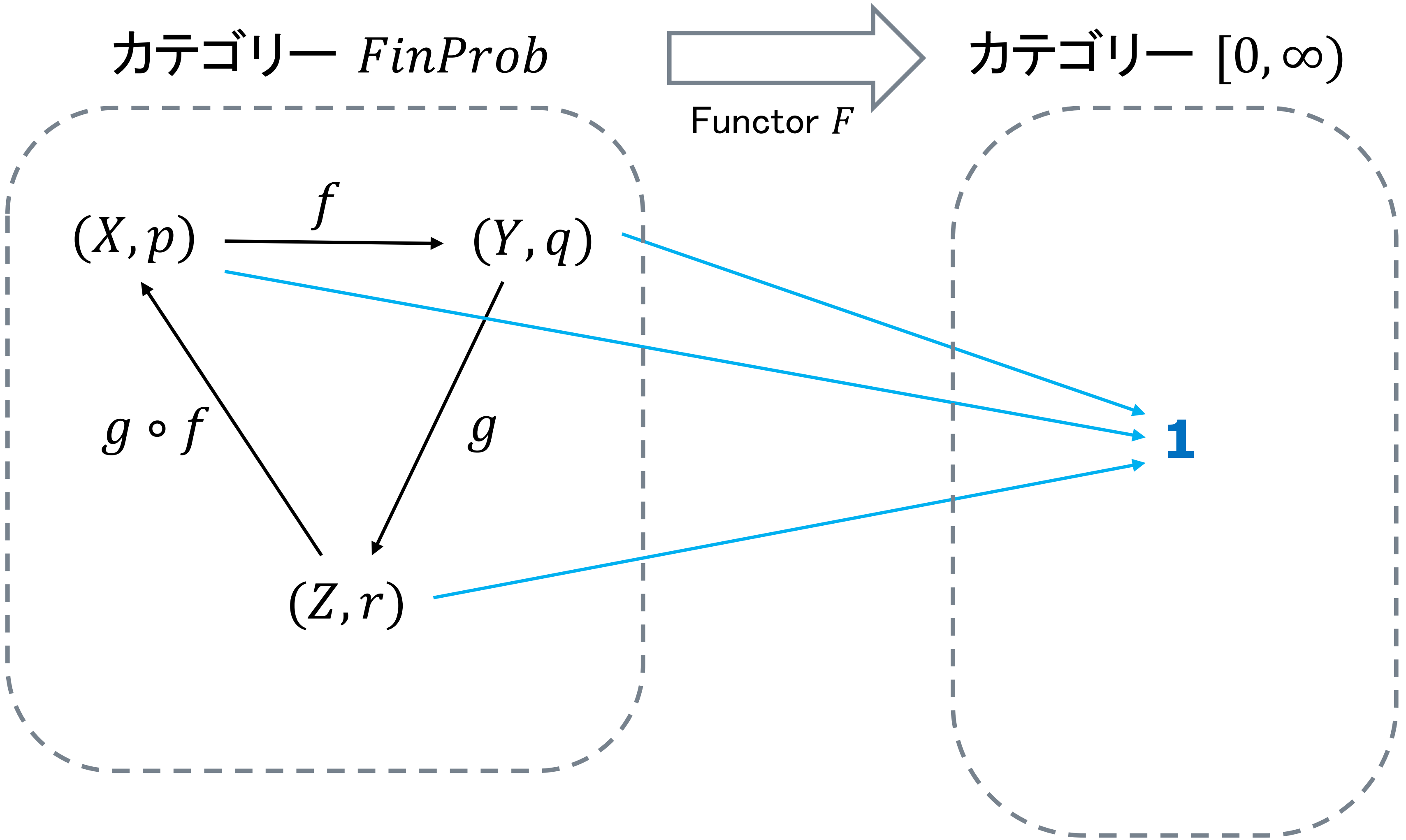
シャノン・エントロピーの
カテゴリー論的解釈の設定

確率分布のカテゴリー$\boldsymbol{FinProb}$と実数のカテゴリー $[\mathbf{0},\infty)$を次のように定義する

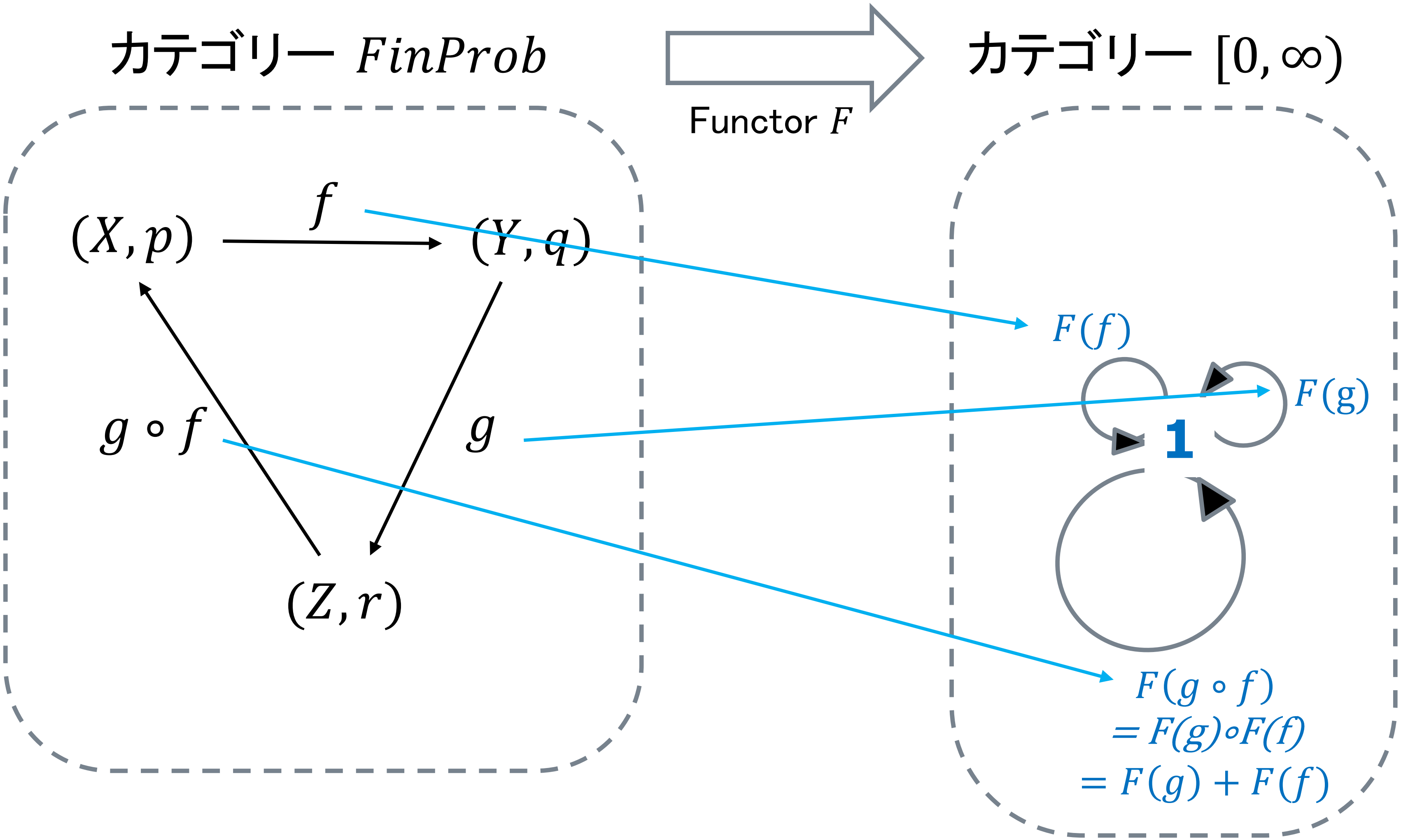
- **FinProb** : 確率分布のカテゴリー
  - オブジェクト : 確率分布 $(\mathrm{X},\mathrm{p}),(\mathrm{Y},\mathrm{q}),(\mathrm{Z},\mathrm{r}),\ldots$
  - 射 : 測度を保存する関数 $f:(X,p)\to(Y,q), g:(Y,q)\to(Z,r),\ldots$
  - 射の合成 : 関数の合成 $f\circ g,\ldots$

- $[\mathbf{0},\infty)$: 実数のカテゴリー
  - オブジェクト : 実数カテゴリーは一つのオブジェクトしか持たない
    これを$*$とか1とかで表す。
  - 射 : $*\to*$
    実数 $r,s,\ldots$ は、射 $*\to*$ で表現される
  - 射の合成 :
    射$r:*\to*$と射$s:*\to*$ の合成は、$r+s:*\to*$として和で定義される

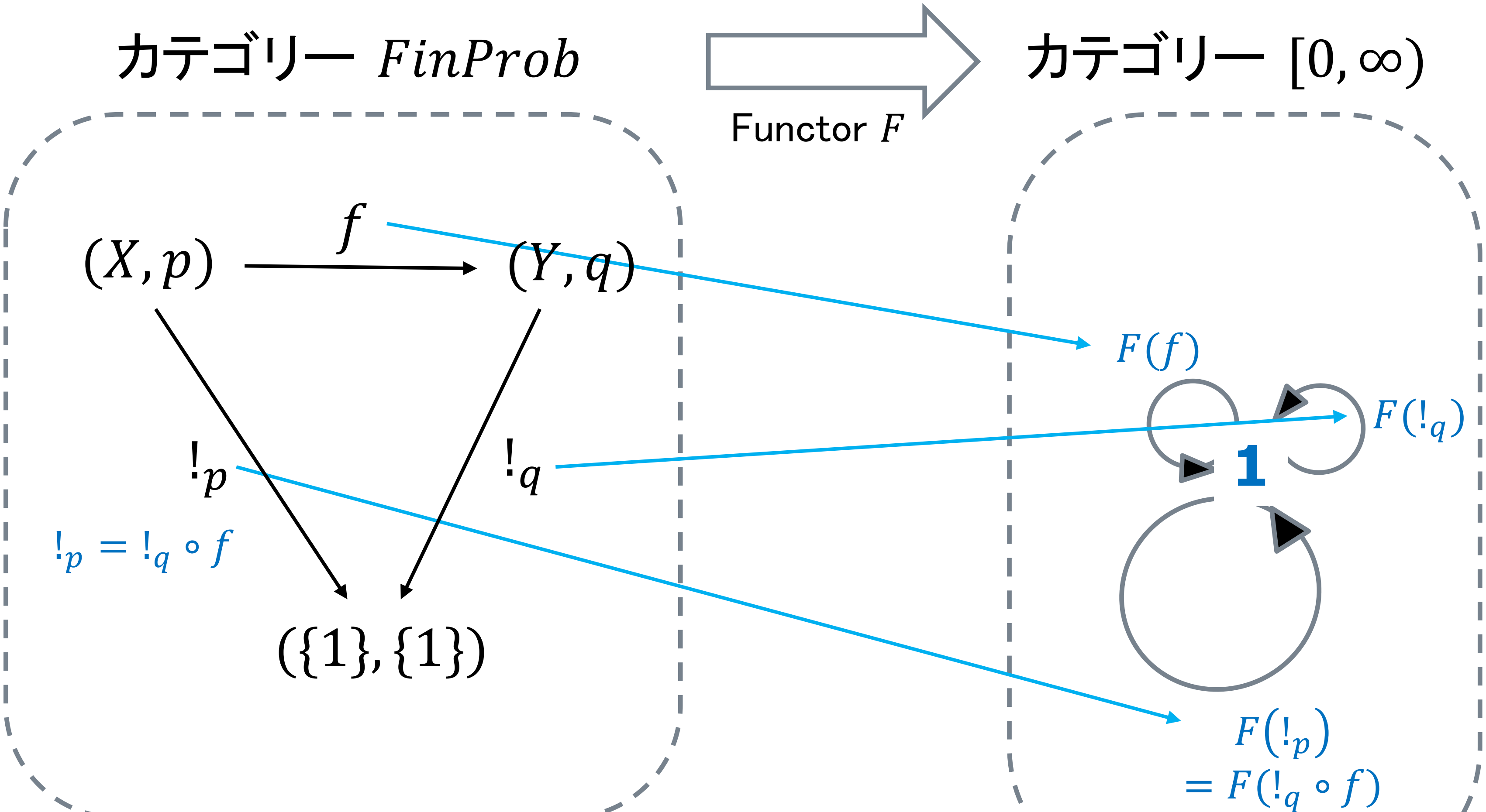
$FinProb$から$[0,\infty)$へのFunctor F
オブジェクトへのFの作用



$FinProb$から$[\mathbf{0},\infty)$へのFunctor F
射への作用



$FinProb$から$[\mathbf{0}, \infty)$へのFunctor F
射 $f, !_p, !_q$への作用



$F(f) = cH(p) - cH(q)$ ⇦ $F(f) = F(!_p) - F(!_q)$ ⇦

$$F(f) = cH(p) - cH(q)$$

カテゴリー $FinProb$

カテゴリー $[0, \infty)$

$$!_p = !_q \circ f$$

$$\begin{tikzcd}
(X, p) \arrow[rr, "f"] \arrow[dr, "!_p"'] & & (Y, q) \arrow[dl, "!_q"] \\
& (\{1\}, \{1\}) &
\end{tikzcd}$$

Functor F

$$F(!_p) = F(!_q \circ f)$$
$$= F(!_q) + F(f)$$

$$F(f) = F(!_p) - F(!_q)$$

$$F(f) = cH(p) - cH(q)$$

実数 $[0,\infty)$をカテゴリーと見るアイデアの起源

実数 $[0,\infty)$をカテゴリーと見るアイデアは、1973年のLawvereの論文 “Metric spaces, generalized logic and closed categories” によるものである。
http://www.tac.mta.ca/tac/reprints/articles/1/tr1.pdf

X	hom-values for X	composition law and identity law for X	domain of composition law for X	domain of identity law for X
metric space	nonnegative real quantities	$\geqq$	sum	zero
category	abstract sets	mapping	cartesian product	one element set

stochastic map

stochastic map $f : X \rightsquigarrow Y$

m個の確率からなる確率分布 (X, p) と n個の確率からなる確率分布 (Y, q) が与えられた時、確率分布 (X, p) を 確率分布 (Y, q) に変換する写像を

stochastic map $f : (X, p) \rightsquigarrow (Y, q)$

という。これを、$f : X \rightsquigarrow Y$ と書いても構わない。

波線の矢印で表されていることに注意。

stochastic map $f : X \rightsquigarrow Y$ を行列で表す

確率分布を列ベクトルの形で書けば、stochastic map $f : (X, p) \rightsquigarrow (Y, q)$ の作用は、m次元のベクトルpをn次元のベクトルqに変換する、$n \times m$の行列で表現される。

$$\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix} = \begin{pmatrix} f_{11} & , f_{12}, \dots, f_{1m} \\ f_{21} & , f_{22}, \dots, f_{2m} \\ \vdots & \vdots \qquad \vdots \\ f_{n1} & , f_{n2}, \dots, f_{nm} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_m \end{pmatrix}$$

ここで、$q_j = \sum_i f_{ji} p_i$ である。

stochastic map $f : X \rightsquigarrow Y$

stochastic map f は、$x \in X,\ y \in Y$ なる全てのx, y のペアに、先の行列の要素 f_{yx} を割り当てる。この f_{yx}を、与えられたxのyの確率と呼ぶ。

$x \in X$を固定して yを動かすと、実数 f_{yx} の集合は、y上の確率分布を形成する。stochastic map f　は、Xの要素にY上の確率分布を対応づける。

この時、次が成り立っている。

$$x \in X, y \in Y\text{なる全ての}x, y\text{について } f_{yx} \geq 0$$

$$x \in X\text{なる全ての}x\text{について} \sum_{y \in Y} f_{yx} = 1$$

こうした性質をstochastic(確率的)という。

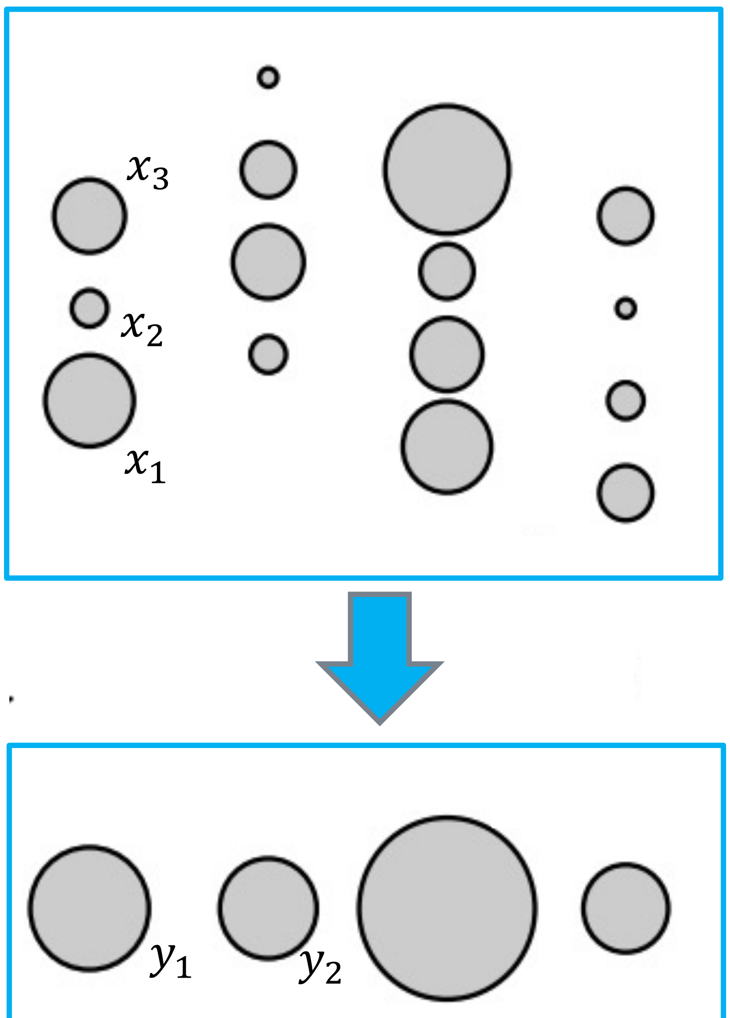
stochastic map $f : X \rightsquigarrow Y$ と stochastic map $g : Y \rightsquigarrow Z$ の合成

$f : X \rightsquigarrow Y$ と $g : Y \rightsquigarrow Z$ の合成 $g \circ f$ を、
f を表す行列と g を表す行列の積で定義する。

$$(g \circ f)_{zx} = \sum_{y \in Y} g_{zy} f_{yx}$$

measure preserving function

測度を保存する関数 f の例
水滴モデル

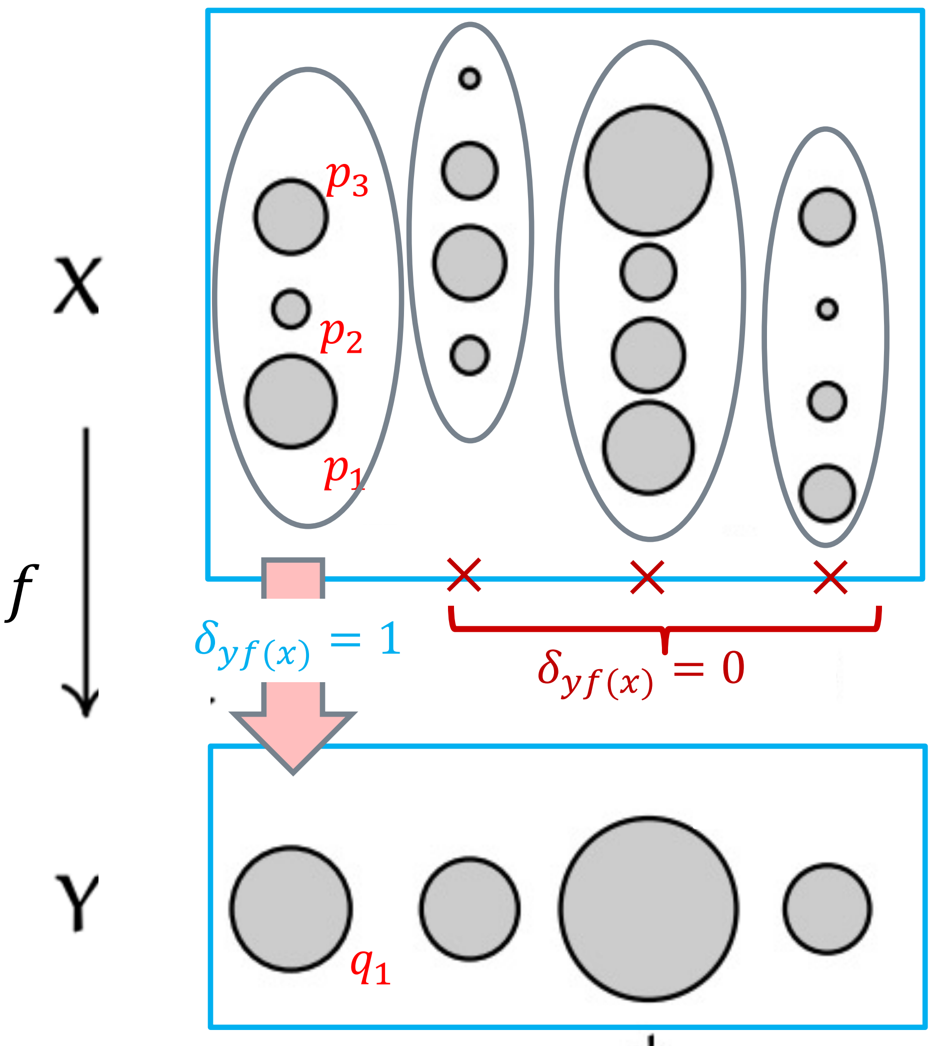


Xは水滴の集まりである。水滴は落下して 豪流して形を変え、水滴の集まりYになったとする。

Xの水滴の体積は、全部で1だとする。この時、Yの水滴の体積も1である。

（要素の体積を全部足して1になるといういう想定は、は、 要素は確率を表し。X,Yは確率分布であることを表現している。）

測度を保存する関数のイメージ



合流前後で体積は変わないことを式で表す

$f(i) = j$である全てのp_iを足したものがq_jとなるので、次の式がなりたつ。

$$q_j = \sum_{x \in X} \delta_{yf(x)} p_x$$

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

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 $f:(X,p)\rightarrow(Y,q)$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。

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

1 からの射 $1 \to X$

ただ一つの要素からなる集合 1

ただ一つの要素からなる集合を 1 で表す。

その一つの要素は、なんであっても構わない。

$$1 = \{*\} = \{\bullet\} = \{1\} = \{2\} = \{\text{丸山}\} = \cdots$$

その一つの要素に何を選ぶかは、以下の議論に影響しない。

Xが集合の場合 $1 \to X$ を考える

$1 \to X$ は、集合 1の上で定義され、集合Xに値を持つ関数を表す。

Xの要素を $\{x_1, x_2, \dots x_n\}$ とする。この時、集合 1の一つの要素 1 と Xの要素 x_i を対応づける関数fには、次のn種類がある。

$$
\begin{gathered}
f_1 : 1 \to x_1 \\
f_2 : 1 \to x_2 \\
\vdots \\
f_n : 1 \to x_n
\end{gathered}
$$

これ以外、定義域が1で値域が Xである関数は存在しない。

$1 \to X$ で特徴づけられる関数はXの要素に一対一に対応する

先に見た定義域が1で値域がXである関数f_i の名前を次のように変えてみる。

$$x_1 : 1 \to x_1$$
$$x_2 : 1 \to x_2$$
$$\vdots$$
$$x_n : 1 \to x_n$$

$1 \to X$ で特徴づけられるn個の関数は、n個のXの要素に一対一に対応することがわかる。

1からの射 $1 \to X$ は、Xの要素を表す

カテゴリー論では、次のように考える。

一つの要素からなるオブジェクト 1 からオブジェクトXへの射 $1 \to X$があったとする。この時、

射 $1 \to X$ は、Xの要素を表す

確率分布の表現

確率分布の表現 1
(X, p)

確率変数 $x_1, x_2, \ldots, x_n$があって、その確率変数がとる確率の値がそれぞれ$p_1, p_2, \ldots, p_n$だとする。この時、この確率分布を次のように表す。

$$((x_1, x_2, \ldots, x_n), (p_1, p_2, \ldots, p_n))$$

p_iは確率なので、$p_1 + p_2 + \cdots + p_n = 1$である。

集合 $X = \{x_1, x_2, \ldots, x_n\}$、$p = \{p_1, p_2, \ldots, p_n\}$ として、この確率分布を次のように表すことができる。

$$(X, p)$$

もっと省略して、「確率分布 X」「確率分布 p」と表すこともある。

確率分布の表現 2
$1 \rightsquigarrow X$

射 $1 \rightsquigarrow X$ は、一つの要素しか持たない確率分布 1から確率分布 Xへの射である。

一つの要素しか持たない確率分布とは、$(\{1\}, \{1\})$という確率分布のことである。前者の$\{1\}$は、確率変数が一つしかないことを表し、後者の$\{1\}$は、その確率が 1 であることを表している。先の表記に従えば、$1 \in \Delta_1$である。

射 $1 \rightsquigarrow X$ は、射 $1 \rightarrow X$の特別の場合と考えることもできる。

射 $1 \rightarrow X$ は、集合Xからその一つの要素を選び出すが、
射 $1 \rightsquigarrow X$ は、確率分布の集合から、一つのXを選び出している。

射 $1 \rightsquigarrow X$ は、Xが確率分布であることを表す。

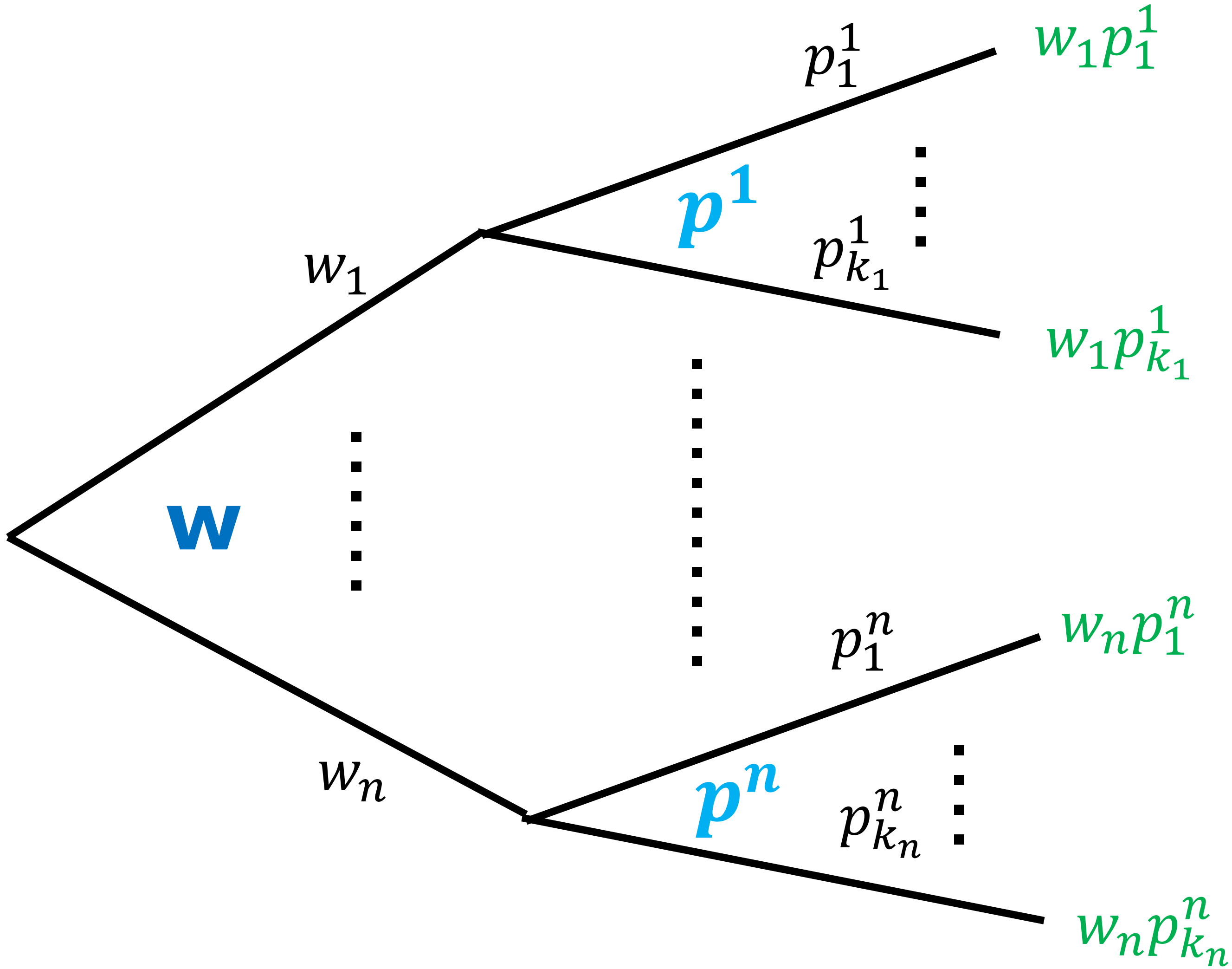
確率分布の表現 3
Δ_n

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。

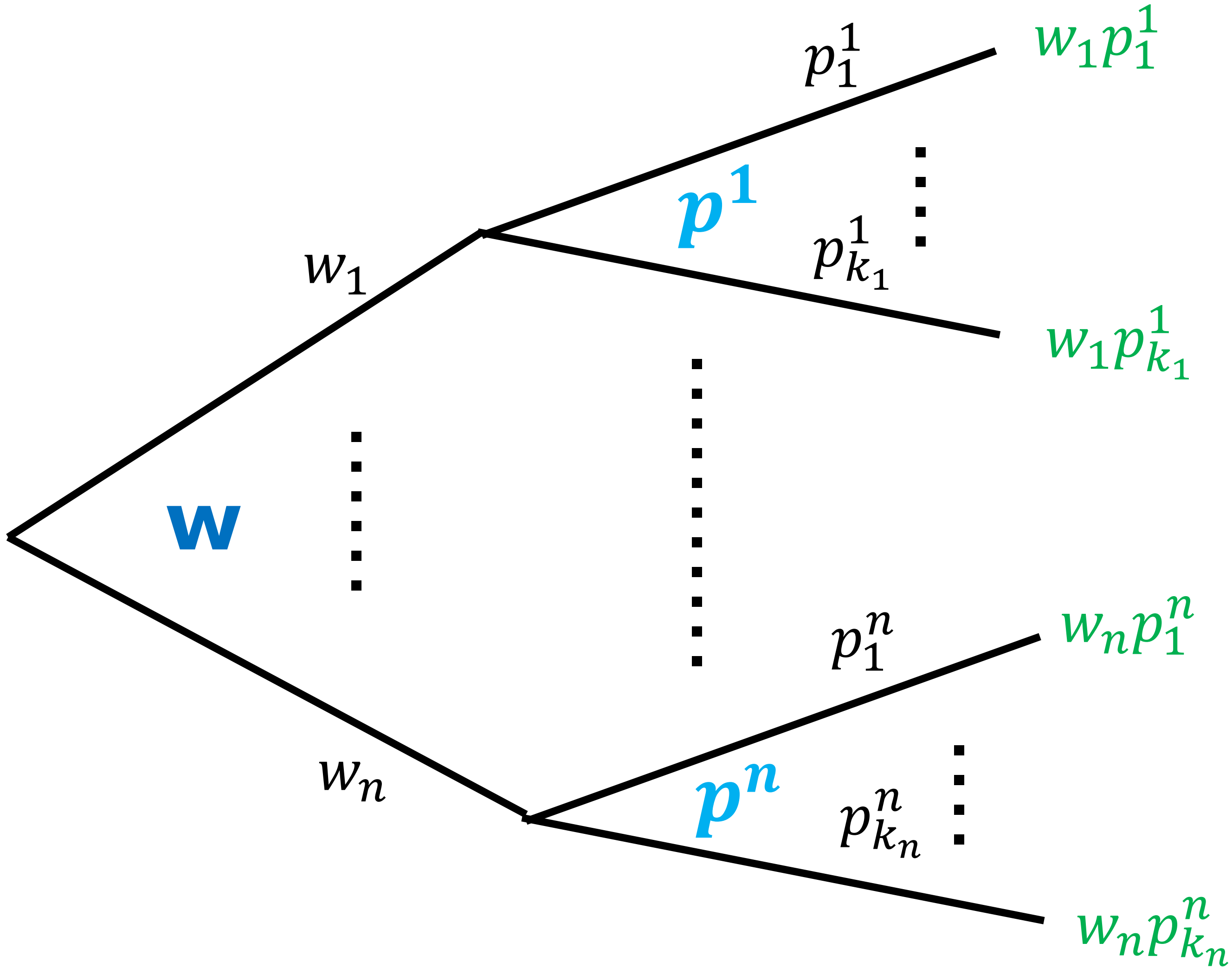
基本的には、n個の確率変数からなる確率分布は、$\{1,2, \dots, n\}$上の確率分布と考えることができる。確率変数の名前は、$\{1,2, \dots, n\}$に、一対一に割り当てればいい。

Δ_nで $\{1,2, \dots, n\}$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



$$\boxed{\begin{array}{c} \boldsymbol{w} \in \Delta_n, \qquad \boldsymbol{p}^1 \in \Delta_{k_1}, \dots, \boldsymbol{p}^n \in \Delta_{k_n} \\ (w_1 p_1^1, \dots, w_1 p_{k_1}^1, \dots, w_n p_1^n, \dots, w_n p_{k_n}^n) \in \Delta_{k_1+k_2+\cdots k_n} \end{array}}$$

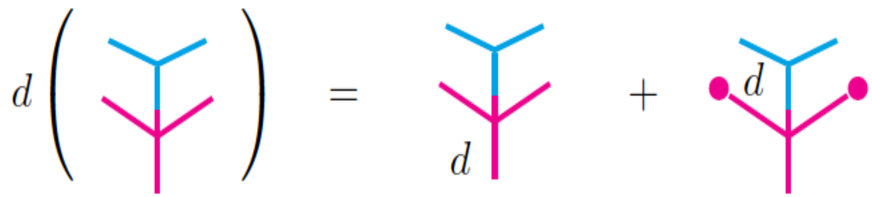


トポロジカルな simplex Δ^nの定義

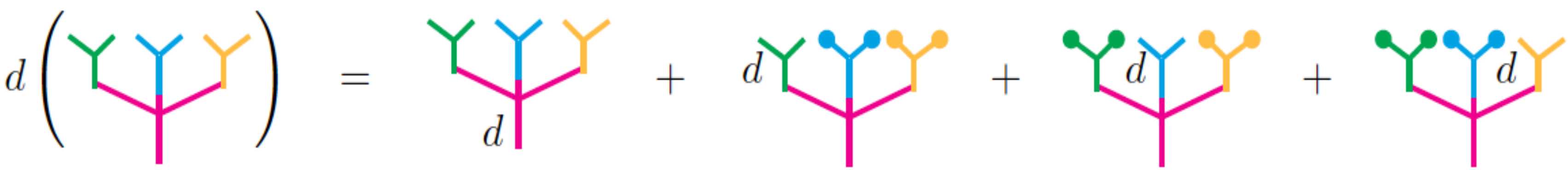
$$\Delta^n := \{(p_0, p_1, \ldots, p_n) \in \mathbb{R}^{n+1} \mid 0 \le p_i \le 1 \text{ and } \sum_{i=0}^{n} p_i = 1\},$$

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

Bradley "Entropy as a Topological Operad Derivation"



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



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

等しい確率分布の表現

stochastic map $q : \mathbf{1} \rightsquigarrow \boldsymbol{X}$ は、X上の確率分布 q を表す。

次の確率分布の表現は、みな等しい。

$$\boxed{\begin{tikzcd} \mathbf{1} \arrow[d, squiggly, "q"'] \\ X \end{tikzcd}} \quad \Longleftrightarrow \quad \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix} \quad \Longleftrightarrow \quad \boxed{\big(X, (q_1, q_2, \dots q_m)\big) = (X, \boldsymbol{q})}$$

on X

$$\Longleftrightarrow \quad \boxed{(X, \boldsymbol{q}) \in \Delta_n}$$

可換な図形

図式が可換

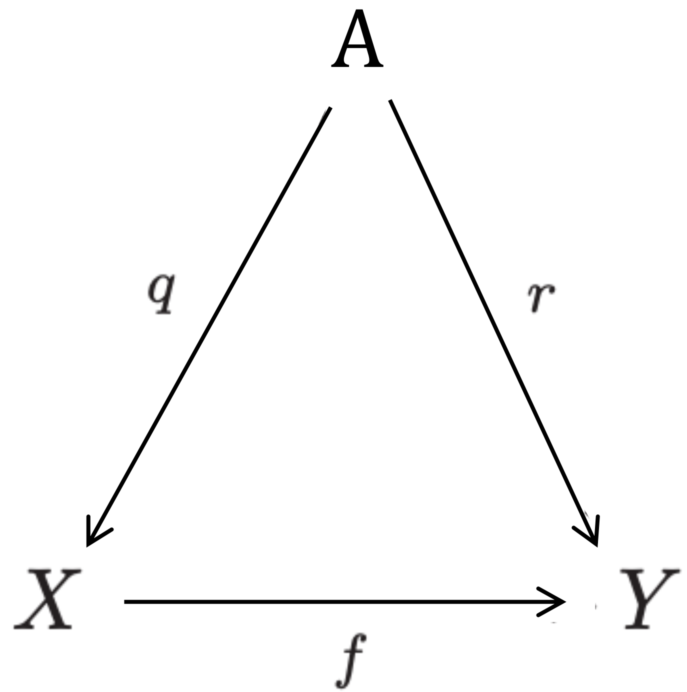
右の図式が「可換」であるというのは、

射 q と射 f の合成が、射 r に等しいことを言う。

すなわち、

$$r = f \circ q$$

が成り立っていることを言う。



stochastic mapの図式が可換だと言う例

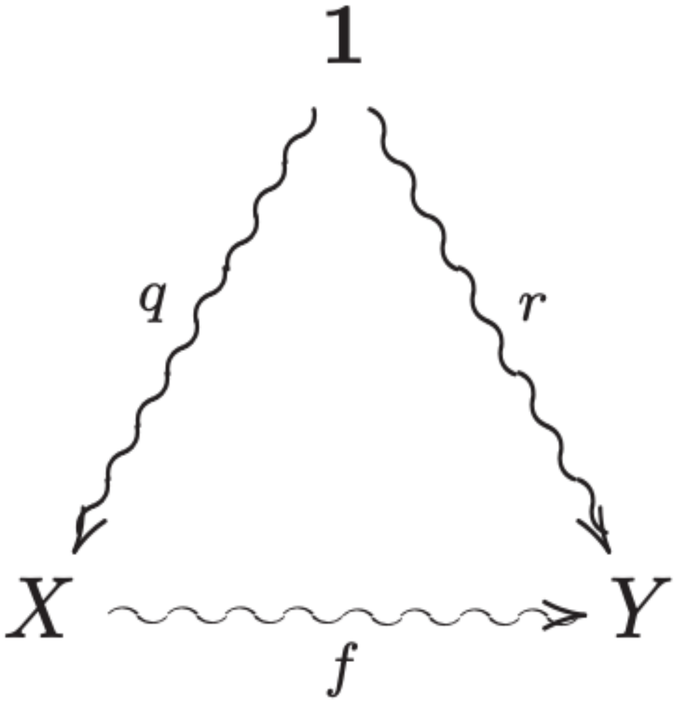
stochastic mapからなる右の図式が可換だということは、

stochastic map q と stochastic map f の合成が、stochastic map r に等しいことを言う。

すなわち、

$$r = f \circ q$$

が成り立っていることを言う。



stochastic mapの図式が可換だと言う例

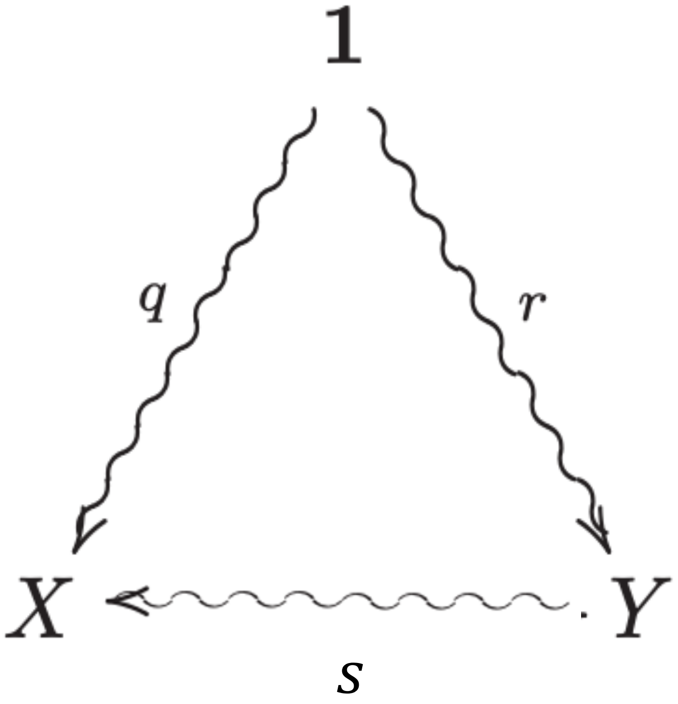
stochastic mapからなる右の図式が可換だということは、

stochastic map r と stochastic map s の合成が、stochastic map q に等しいことを言う。

すなわち、

$$q = s \circ r$$

が成り立っていることを言う。



stochastic mapとmeasure preserving functionの図式が可換だと言う例

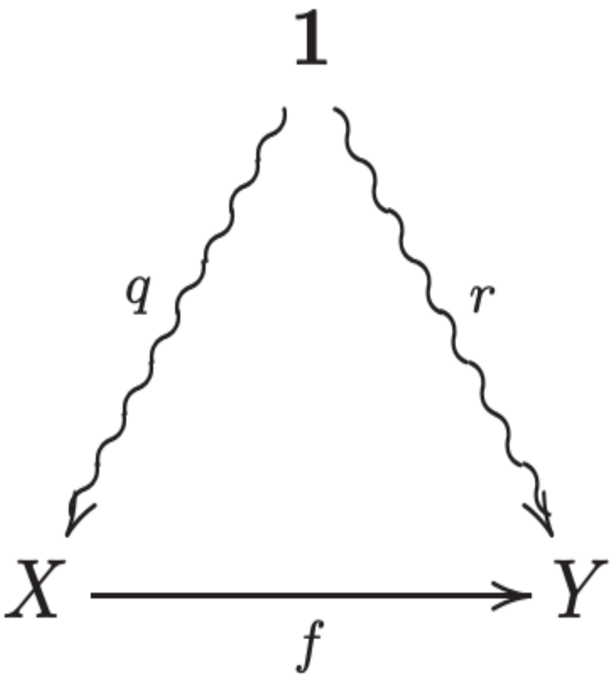
stochastic mapとmeasure preserving functionからなる右の図式が可換だということは、

stochastic map q と measure preserving function f の合成が、stochastic map r に等しいことを言う。

すなわち、

$$r = f \circ q$$

が成り立っていることを言う。



measure preserving function

次の図式が可換だということは？

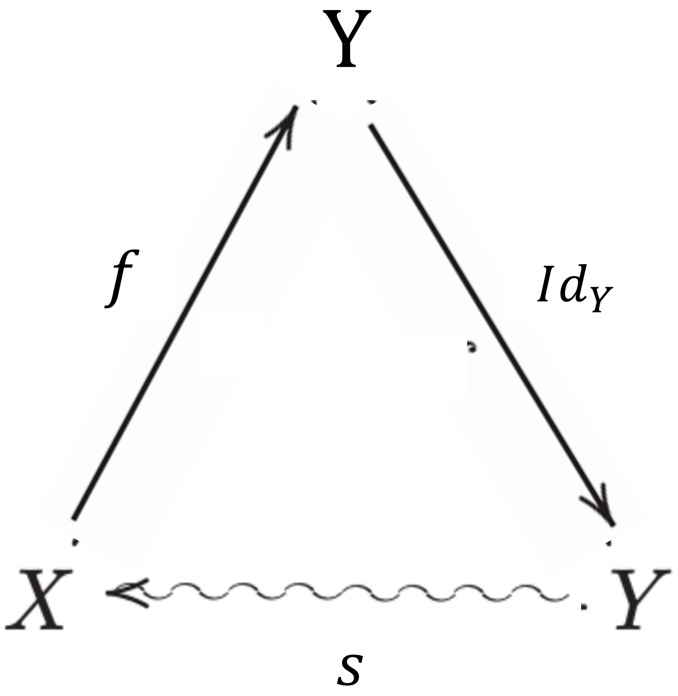
右の図式が可換であるとは、

stochastic map s と measure preserving function f の合成が、同一射 Id_Yに等しいことを言う。

すなわち、

$$Id_Y = f \circ s$$

が成り立っていることを言う、



可換な図形が表現するもの

stochastic mapの 次の可換図式 $r = f \circ q$ の意味

$q\colon 1 \rightsquigarrow X$は、
X上の確率分布**q**を与え、
$r\colon 1 \rightsquigarrow X$は、
Y上の確率分布**r**を与える。

map fの行列をf_{yx}とすると
$r = f \circ q$より、

$$\boxed{r_y = \sum_{x \in X} f_{yx}\, q_x}$$

$$\begin{tikzcd} & \mathbf{1} \arrow[dl, squiggly, "q"'] \arrow[dr, squiggly, "r"] & \\ X \arrow[rr, squiggly, "f"'] & & Y \end{tikzcd}$$

$$\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix} \qquad \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

q on X

r on Y

この可換図式の特殊な場合

$$r_y = \sum_{x \in X} f_{yx}\, q_x$$

この可換図式の特殊な場合として、f の行列が次の形で表される場合を考えよう。

$$\boxed{f_{yx} = \delta_{yf(x)}}$$

δはクロネッカーのデルタ

$$\begin{tikzcd} & \mathbf{1} \arrow[dl, squiggly, "q"'] \arrow[dr, squiggly, "r"] & \\ X \arrow[rr, squiggly, "f"'] & & Y \end{tikzcd}$$

$$\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix} \qquad \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

q on X

r on Y

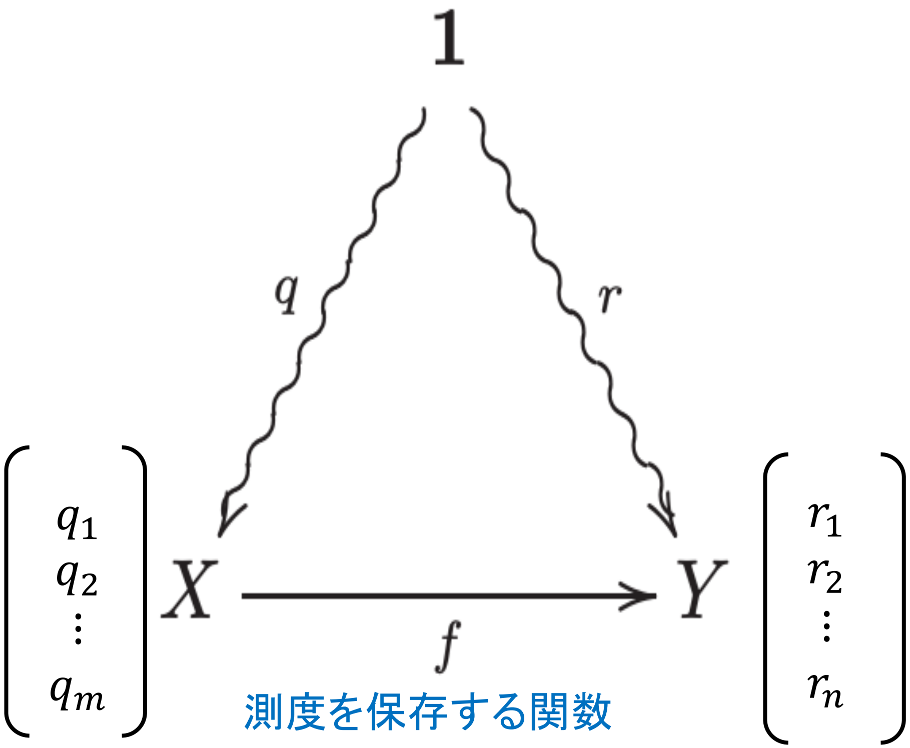
先の可換図式の特殊な場合
測度を保存する関数

$$
\begin{aligned}
r_y &= \sum_{x \in X} f_{yx}\, q_x \\
&= \sum_{x \in X} \delta_{y f(x)}\, q_x \\
&= \sum_{x : f(x) = y} q_x
\end{aligned}
$$

こうしたfを

測度を保存する関数

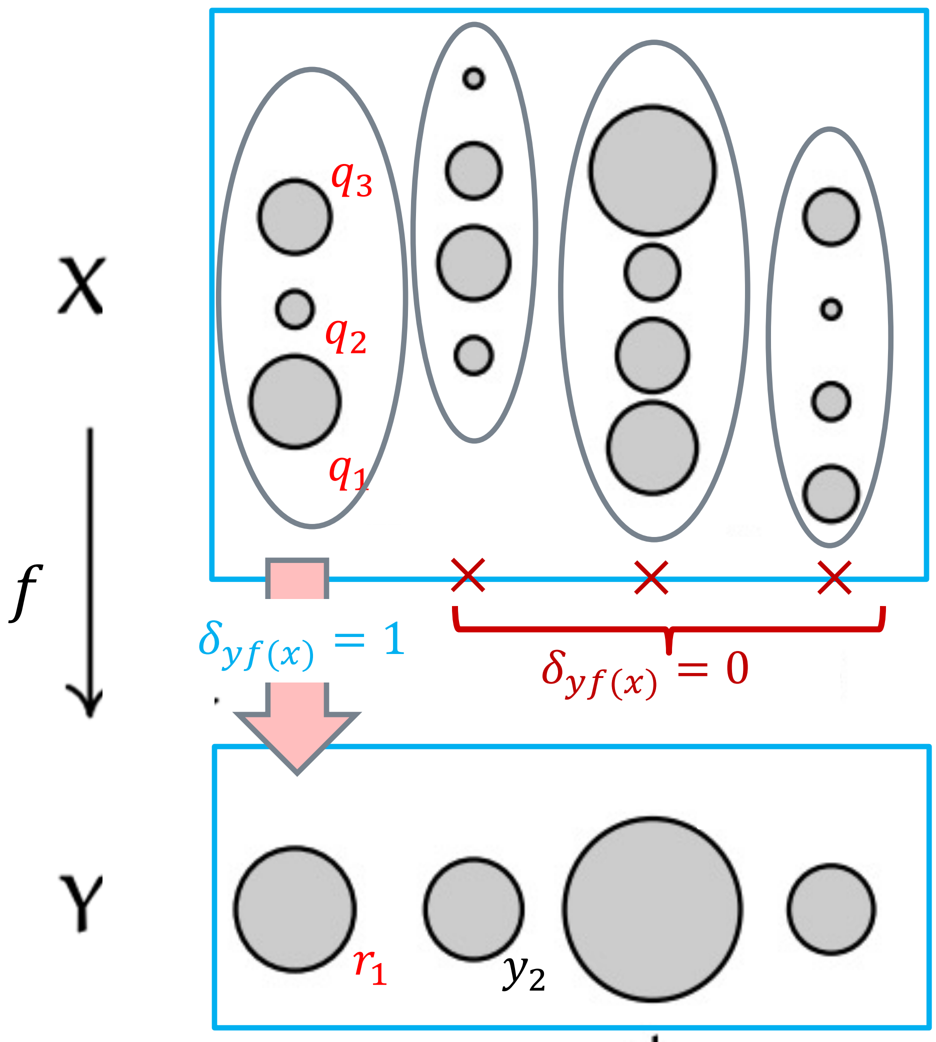
という。



X上の
確率分布q

Y上の
確率分布r

測度を保存する関数のイメージ



合流前後で体積は変わないことを式で表す

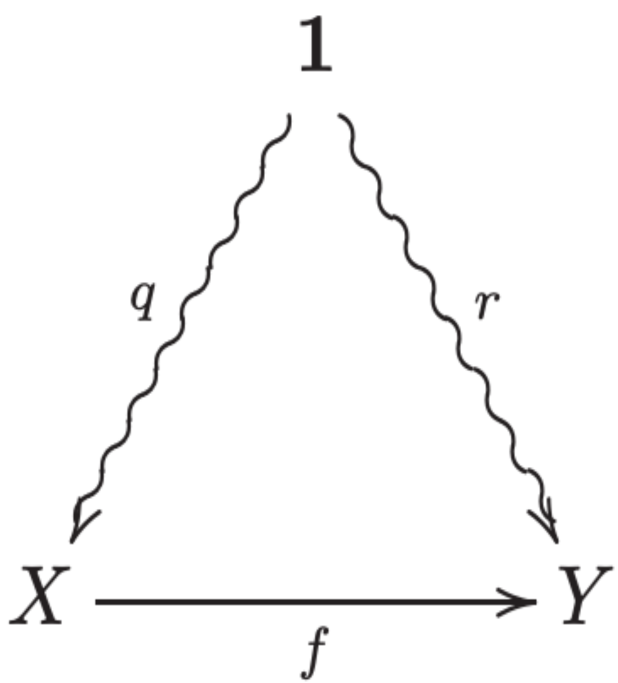
$f(i) = j$である全てのx_iを足したものがy_jとなるので、次の式がなりたつ。

$$r_j = \sum_{x \in X} \delta_{yf(x)} q_x$$

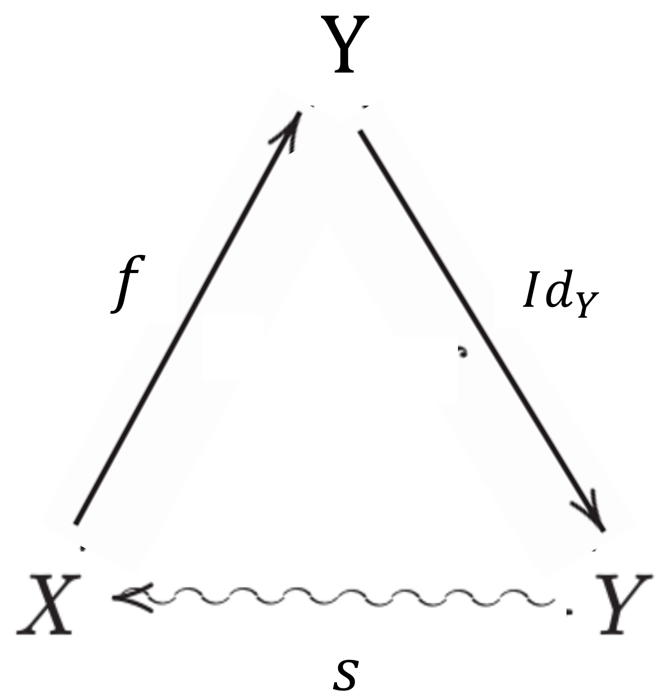
$$= \sum_{i:f(i)=j} q_i$$

相対エントロピーの
カテゴリー論的解釈の設定

次の二つの図式が可換な場合を考える



measure preserving function

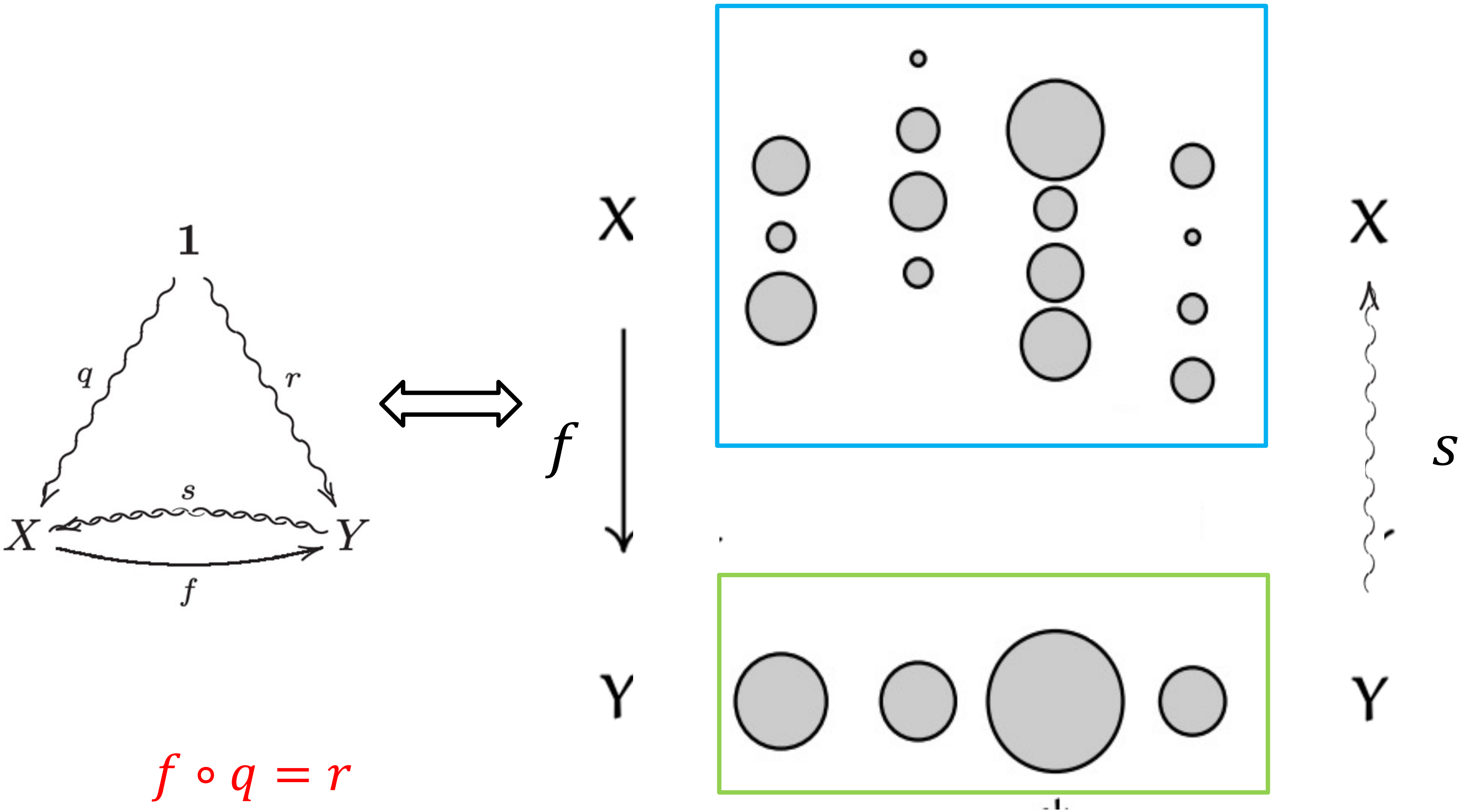


$$f \circ q = r$$

$$f \circ s = Id_Y$$

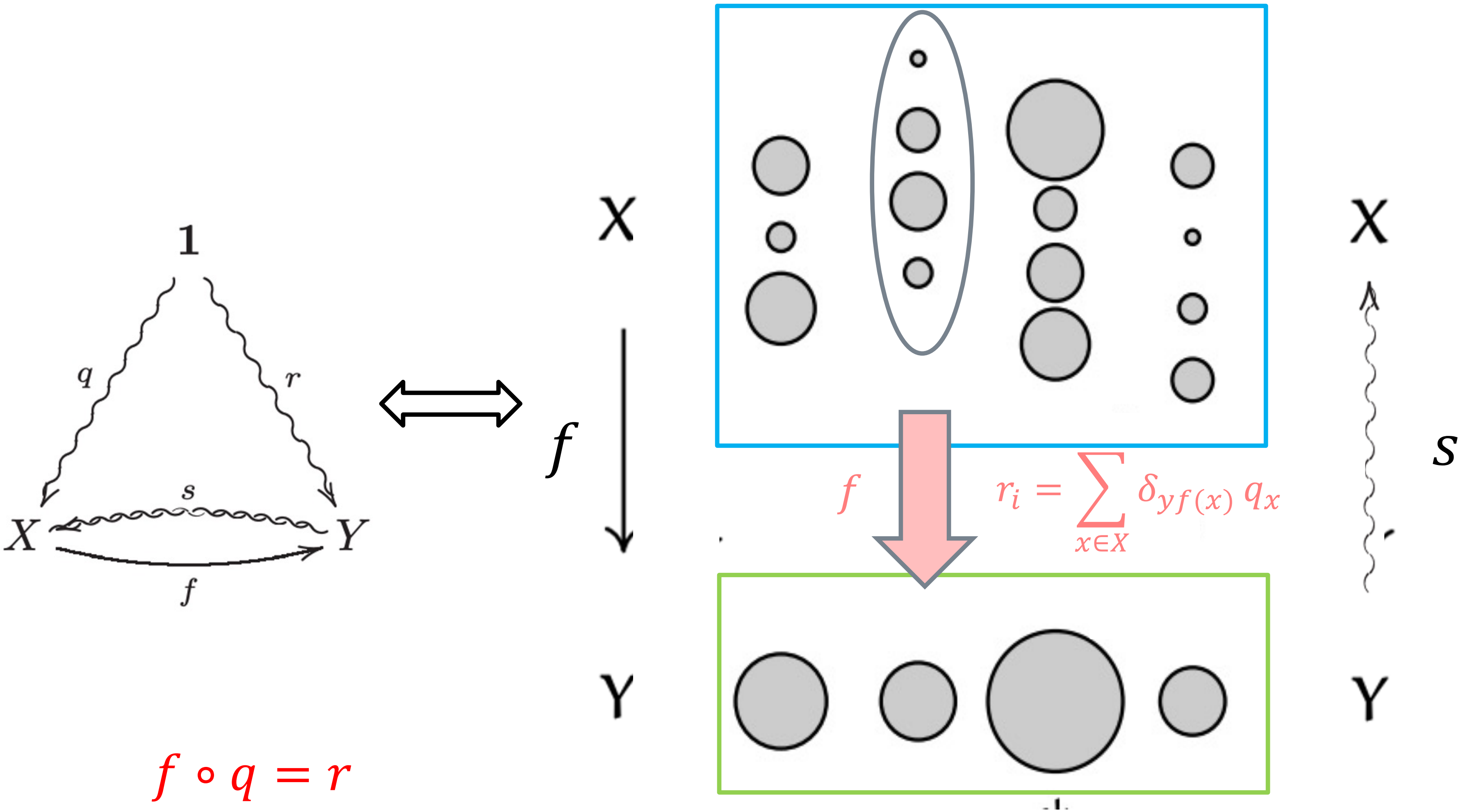
$$
\begin{tikzcd}
 & \mathbf{1} \arrow[dl, squiggly, "q"'] \arrow[dr, squiggly, "r"] & \\
X \arrow[rr, bend right, "f"'] & & Y \arrow[ll, squiggly, bend right, "s"']
\end{tikzcd}
$$

$$
\begin{aligned}
f \circ q &= r \\
f \circ s &= 1_Y
\end{aligned}
$$



$$f \circ q = r$$

$$f \circ s = 1_Y$$

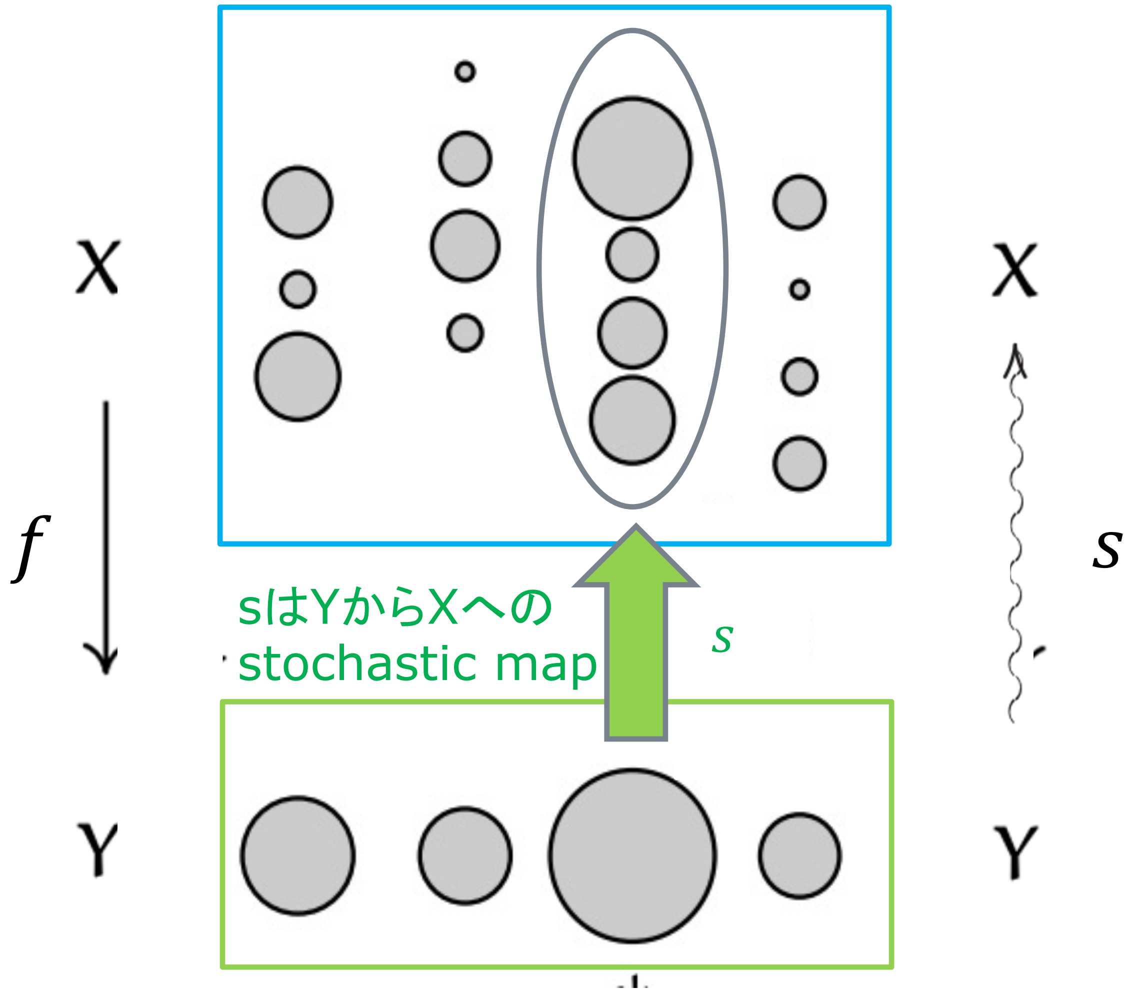


$f \circ q = r$

$f \circ s = 1_Y$

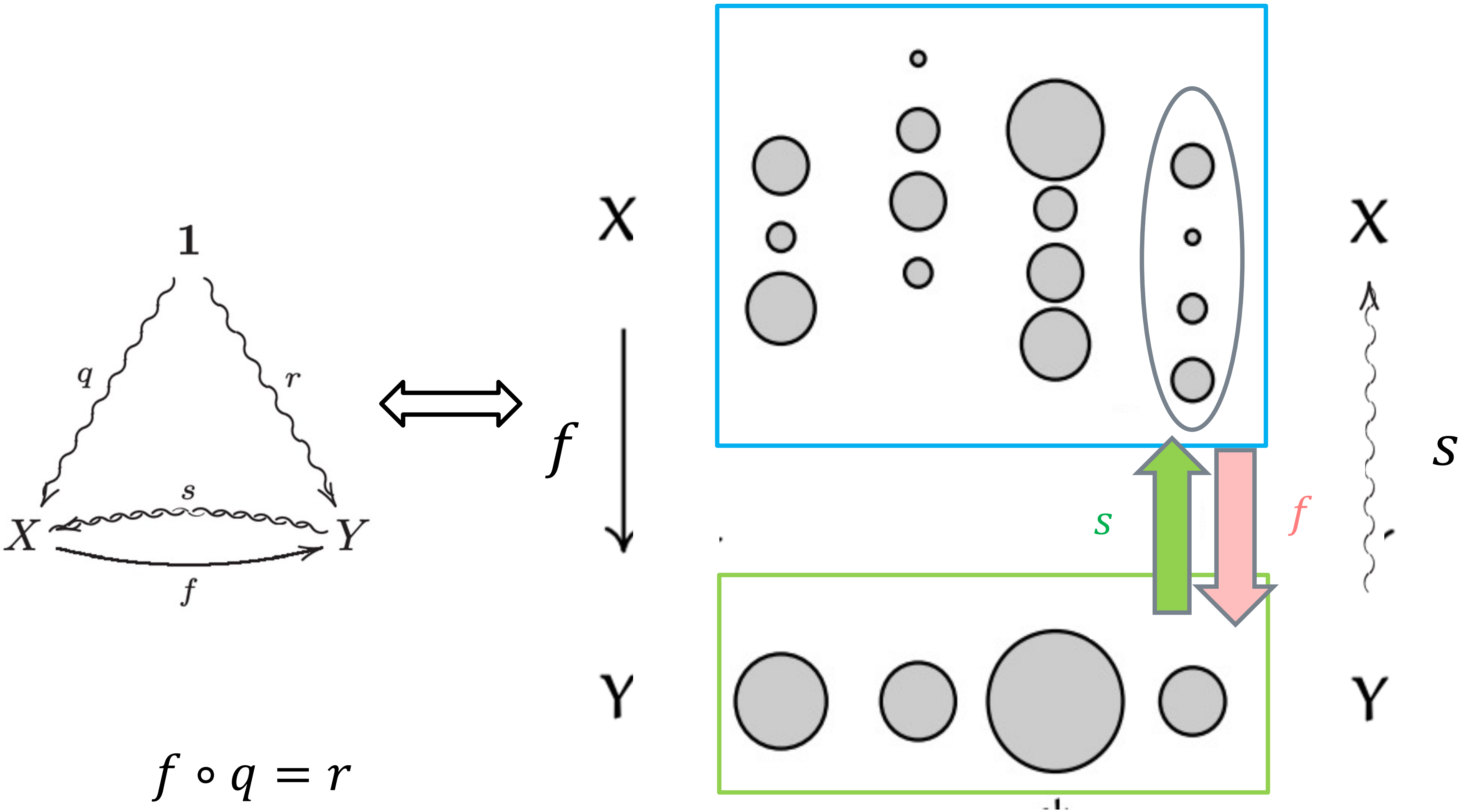
$$
\begin{tikzcd}
 & \mathbf{1} \arrow[dl, squiggly, "q"'] \arrow[dr, squiggly, "r"] & \\
X \arrow[rr, bend right, "f"'] & & Y \arrow[ll, squiggly, "s"']
\end{tikzcd}
$$

$\Longleftrightarrow$



$$f \circ q = r$$

$$f \circ s = 1_Y$$

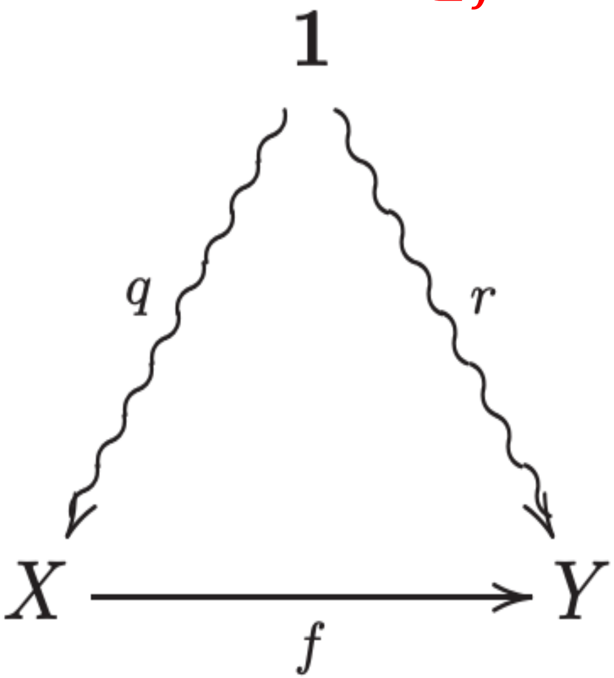


$$f \circ q = r$$

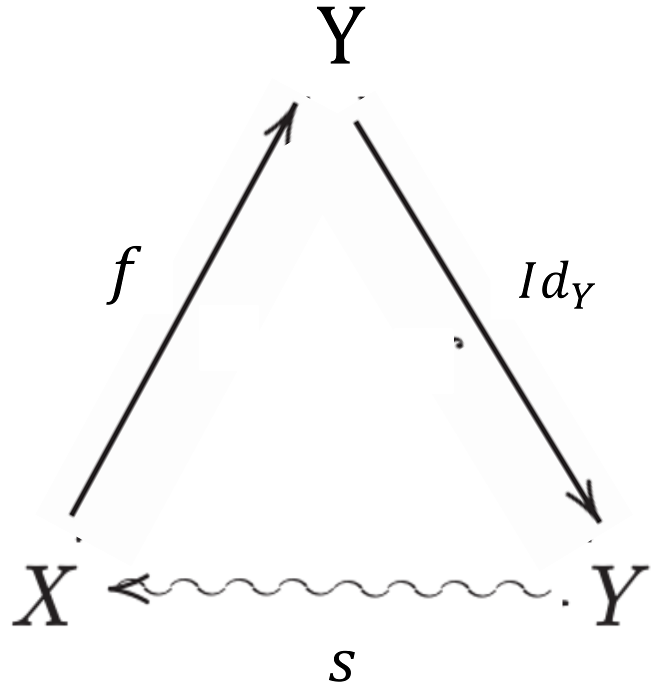
$$f \circ s = 1_Y$$

二つの図式が可換
disintegration

sをfの$disintegration$と呼ぶ



measure preserving function



$f \circ q = r$

$f \circ s = Id_Y$

第三部

相対エントロピーの
カテゴリー論的解釈

第三部
相対エントロピーのカテゴリー論的解釈

1. 相対エントロピー
2. stochastic map
3. stochastic mapとFinProb
4. Finstatと「仮説」
5. Functorとしての相対エントロピー

相対エントロピー

セッションの概要

この連続セッションでは、相対エントロピーへのカテゴリー論的アプローチとして、BaezとFritzの2014年の論文 ”A Bayesian Characterization of Relative Entropy” を紹介する。
https://arxiv.org/abs/1402.3067v2

この論文は、先に紹介した、エントロピー論に、”Entropy As a Functor” という視点を初めて導入した、Baez, Fritz, Leinster の 2011年の論文 “A Characterization of Entropy in Terms of Information Loss” の続編である。
https://arxiv.org/abs/1106.1791v3

Entropy as a Functor

2011年の論文は、シャノン・エントロピーの「差」としての「情報の損失」に注目する。

確率分布のカテゴリー **FinProb** から実数のカテゴリー $[\mathbf{0}, \infty)$ へのFunctor Fが、ある性質を満たすとき、カテゴリーFinProb上の測度を保存する射fの $F(f)$が、一意にシャノン・エントロピー（の「差」の定数倍）を決定することを示すことが出来る。

$$F\colon \; \boldsymbol{FinProb} \;\to\; [\mathbf{0}, \infty)$$
$$\left((X, p) \xrightarrow[f]{} (Y, q)\right) \;\mapsto\; \mathrm{c}(\mathrm{S}(\mathrm{p}) - \mathrm{S}(\mathrm{q}))$$

ひるがえって、このFunctorの性質を、シャノン・エントロピーの性質とみなすことができるという議論である。

Entropy as a Functor 2

2014年の論文の構成は、2011年の論文よりすこし複雑である。確率分布のカテゴリー **FinProb** に代わって新しいカテゴリー **FinStat**が導入される。

FinStatは、測度を保存する射fに加えて、「仮説」を表すものと解釈されるもう一つの射 r を持つ。

$[\mathbf{0}, \infty]$を、一つのオブジェクトをもち、∞ を含非負の実数を射とし、射の合成は和で定義されているるカテゴリーとしよう。

Entropy as a Functor 2

FinStatからカテゴリー $[\mathbf{0},\infty]$へのFunctor RE が、ある性質を満たすとき、このFunctorが、相対エントロピー を決定することを示すことが出来る。

$$\mathrm{RE}\colon \mathtt{FinStat} \to [0,\infty]$$

$$\left((X,q) \underset{f}{\overset{s}{\leftrightarrows}} (Y,r) \right) \mapsto S(q, s\circ r)$$

相対エントロピー

相対エントロピー

pに対するqの「相対エントロピー」$\mathrm{H_{rel}(q, p)}$を、次の式で定義する。

$$H_{rel}(q,p) = \sum_{x} q(x) log \frac{q(x)}{p(x)}$$

$H_{rel}(q,p)$ =0となるのは、$q = p$ の場合だけであることはすぐわかる。

$H_{rel}(q,p)$を $H(q(x)||p(x))$ と表すこともある。

相対エントロピーの Bayesian的解釈

「相対エントロピー」 $\mathrm{H_{rel}(q, p)}$ に次のような解釈を与えよう。

事前に知っていた（多分、それは正確な知識ではないかもしれないので「仮説」といってもいい）確率分布をp_i としよう。実際に、観測して得られた新しい確率分布 q_i とする。

こうしたアプローチは、Baysianのものである。

「相対エントロピー」というのは、アプリオリな「シャノンのエントロピー」を、Baysianの考え方で、相対化したエントロピーと考えることができる。

認識と学習のBaysian的解釈

エントロピー＝情報量のこの相対的な解釈は、人間の認識で得られる情報量の解釈には、とても向いている。認識や学習のモデルを、この情報量を使って解釈できる。

例えば、先の$H_{rel}(q,p)$ ＝0の場合の解釈では、仮説pと実験結果qが一致した場合には、実験で得られた情報量は0 だと考えればいい。

認識と学習のBaysian的解釈

認識の順番をt で表してみよう。

先の例では、事前の仮説 $p(t-1)$が、実験によって、事後に $q(t)$ に置き換わるのだが、この実験で得られた情報は、$H_{rel}\big(q(t), p(t-1)\big)$ で表される。

ここで、$q(t)$ を新たな$p(t)$ として、$H_{rel}\big(q(t+1), p(t)\big)$ を考える操作を繰り返すことができる。

これは、「認識の発展」のモデルと考えることができる。この「認識の発展」は、$H_{rel}(q, p) = 0$になるときに終わる。

認識と学習のBaysian的解釈

逆に、もしも、最初から正しい分布q を、何らかの方法で我々が知っていて、実験 $p(t)$を繰り返すのなら、$H_{rel}\big(q, p(t)\big)$は、実測値$p(t)$から、正しい答えq に至るために「学習しなければいけない情報量」を表すことになる。（q は、「常に正しい」と仮定しているので、それは時間には依存しない。qは、tを含まないことに注意。）

ここでは、$H_{rel}\big(q, p(t)\big)$ =0 は、「もはや、学習すべき情報が残されていない」ことを意味して、その状態で、学習は終わる。

stochastic map

Entropy as a Functor 2

FinStatからカテゴリー $[\mathbf{0}, \infty\,]$へのFunctor RE が、ある性質を満たすとき、このFunctorが、相対エントロピー を決定することを示すことが出来る。

$$
\begin{aligned}
\mathrm{RE}\colon \mathtt{FinStat} &\to [0,\infty] \\
\left(\begin{tikzcd} (X,q) \arrow[r, bend right, "f"'] & (Y,r) \arrow[l, squiggly, bend right, "s"'] \end{tikzcd} \right) &\mapsto S(q, s \circ r)
\end{aligned}
$$

相対エントロピー

前回、「セッションの概要」で見た図

Entropy as a Functor 2

FinStatからカテゴリー $[\mathbf{0},\infty]$へのFunctor RE が、ある性質を満たすとき、このFunctorが、相対エントロピー を決定することを示すことが出来る。

$$\mathrm{RE} : \textbf{FinStat} \to [0,\infty]$$

$$\left(\begin{tikzcd} (X,q) \arrow[r, bend right, "f"'] & (Y,r) \arrow[l, squiggly, bend right, "s"'] \end{tikzcd} \right) \mapsto S(q, s\circ r)$$

相対エントロピー

カテゴリー **FinStat**

Entropy as a Functor 2

FinStatからカテゴリー $[\mathbf{0},\infty]$へのFunctor RE が、ある性質を満たすとき、このFunctorが、相対エントロピー を決定することを示すことが出来る。

$$\mathrm{RE}\colon \textbf{FinStat} \to [0,\infty]$$

$$\left(\begin{tikzcd} (X,q) \arrow[r, bend right, "f"'] & (Y,r) \arrow[l, squiggly, bend right, "s"'] \end{tikzcd} \right) \mapsto S(q, s \circ r)$$

相対エントロピー

カテゴリー FinStat の射

$f\colon (X,q) \to (Y,r)$

$s\colon (Y,r) \rightsquigarrow (X,q)$

FinStochと stochastic map $f : X \rightsquigarrow Y$

この論文の主役は、カテゴリー **FinStat**なのだが、それについて語る前に、有限集合の(確率論的)カテゴリー**FinStoch**とその上の射の**stochastic maps**（確率写像）について触れておく。

stochastic map f : X ⇝ Y は、普通の関数とは異なっている。普通の関数なら、Xの要素のそれぞれに対して、Yの要素を一つユニークに割り当てるのだが、stochastic map f は、Xの要素のそれぞれに対して、Y上の確率分布を割り当てる。

だから、f(x)は、Yの特定の要素ではなく、異なった値を取る確率を持つ。これが、stochastic map を表すのに、ギザギザ付きの矢印を使い理由である。

stochastic map $f : X \rightsquigarrow Y$

有限集合 X と Y が与えられた時、
stochastic map $f : X \rightsquigarrow Y$ は、$x \in X,\ y \in Y$ なる全てのx, y のペアに、実数 f_{yx} を割り当てる。

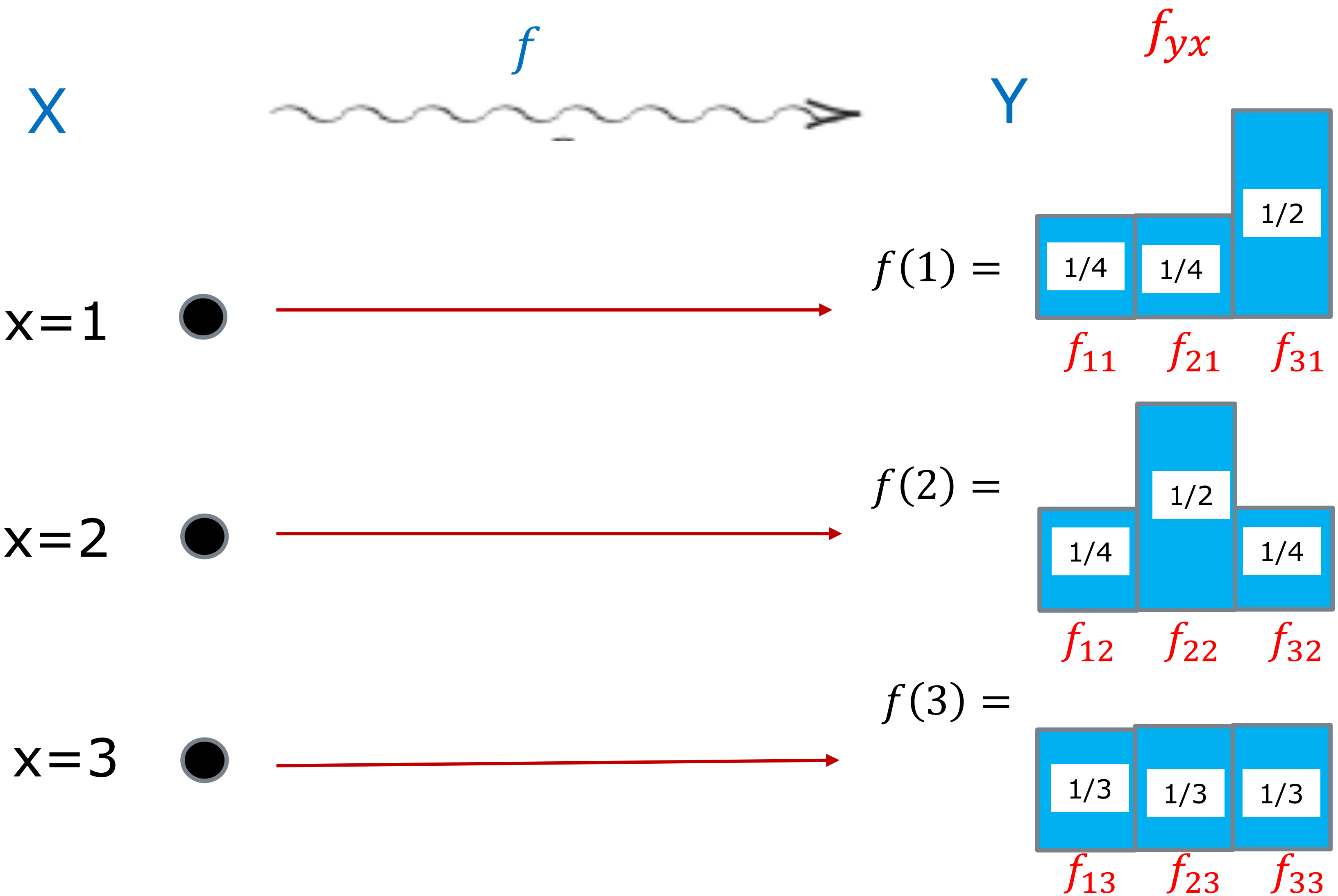
stochastic map f は、Xの要素のそれぞれに対して、Y上の確率分布を割り当てるというのは、次のようなことだ。
$x \in X$を固定すると、実数 f_{yx} は、y上の確率分布を形成する。この f_{yx} を、与えられたxのyの確率と呼ぶ。この時、次が成り立っている。

$$x \in X, y \in Y\text{なる全ての}x, y\text{について } f_{yx} \geq 0$$

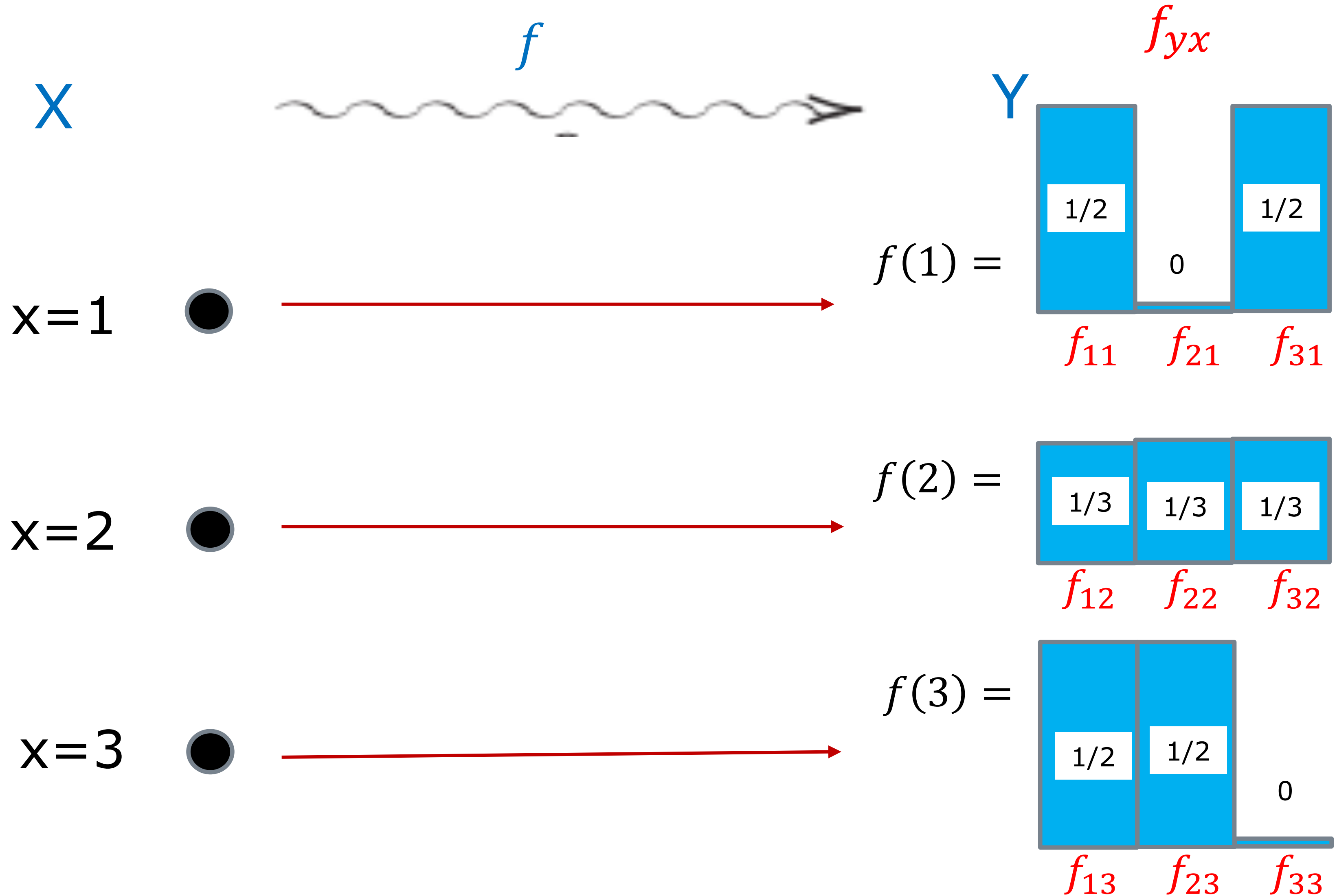
$$x \in X\text{なる全ての}x\text{について} \sum_{y \in Y} f_{yx} = 1$$

こうした性質をstochastic(確率的)という。

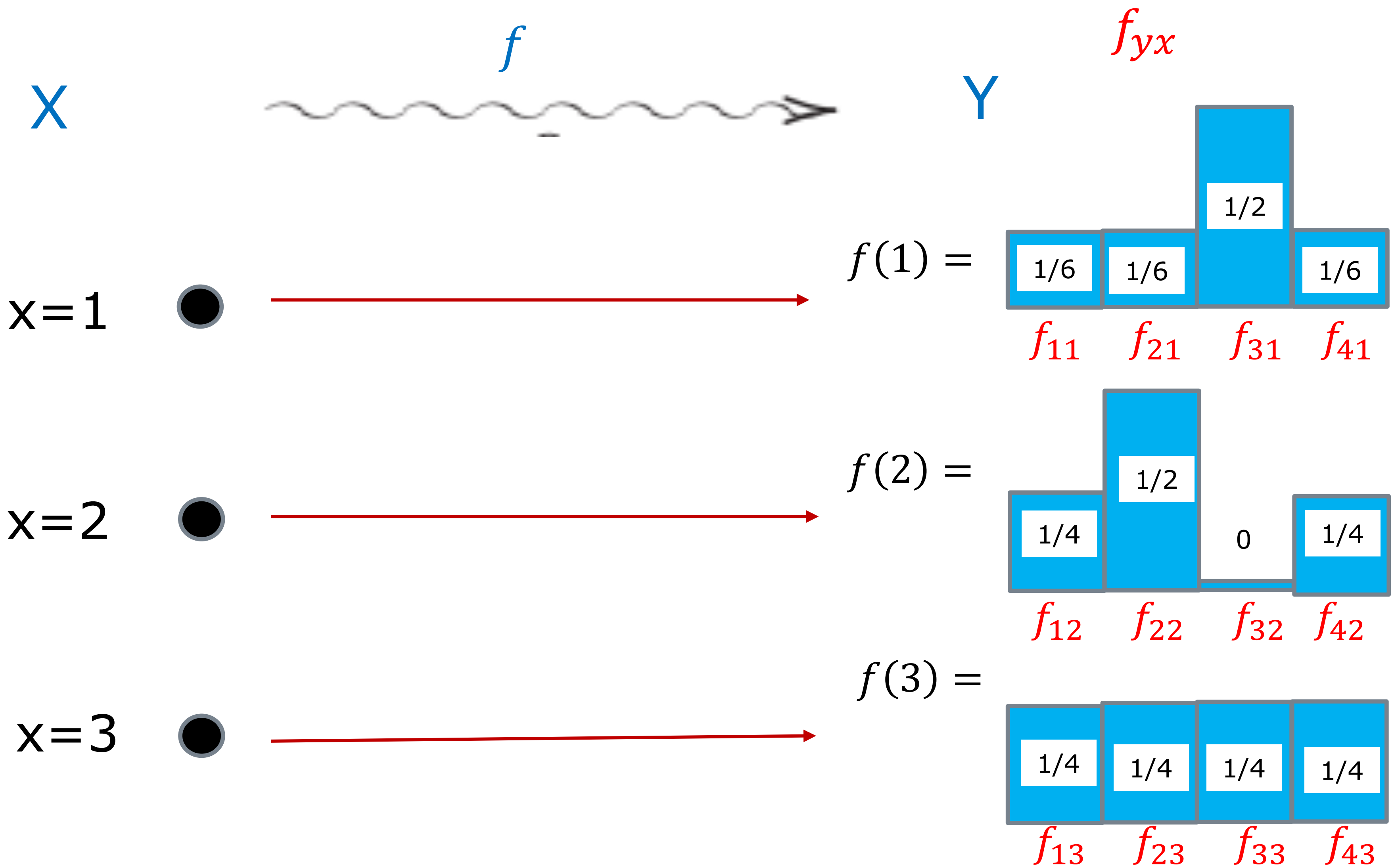
stochastic map $f : X \rightsquigarrow Y$ の例 1



stochastic map $f : X \rightsquigarrow Y$ の例 2



stochastic map $f: X \rightsquigarrow Y$ の例 3



stochastic map $f : X \rightsquigarrow Y$ を行列で表す

有限集合 X と Y が与えられた時、
stochastic map $f : X \rightsquigarrow Y$ は、$x \in X$, $y \in Y$ なる全てのx, y のペアに、実数 f_{yx} を割り当てる。見やすいように、f_{yx}を$f(y,x)$と書くことにしよう。

Yがn個の要素からなり、Xがmこの要素からなる時、
stochastic map $f : X \rightsquigarrow Y$ を次のような行列で表現することができる。

$$\begin{pmatrix} f(1,1) & ,f(1,2) & \dots, f(1,m) \\ f(2,1) & ,f(2,2) & \dots, f(2,m) \\ \vdots & \vdots & \vdots \\ f(n,1) & ,f(n,2) & \dots, f(n,m) \end{pmatrix}$$

stochastic map $f : X \rightsquigarrow Y$ を行列で表す 例 1

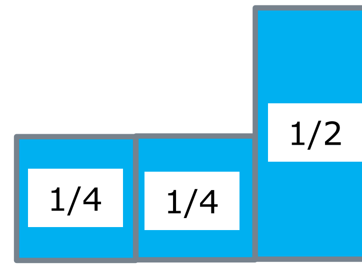
$$X \overset{f}{\rightsquigarrow} Y$$

f_{yx}

$$\begin{array}{cc} & \text{x=1}, \quad \text{x=2}, \quad \text{x=3} \\ \begin{array}{c} \text{y=1} \\ \text{y=2} \\ \text{y=3} \end{array} & \begin{pmatrix} 1/4, & 1/4, & 1/3 \\ 1/4, & 1/2, & 1/3 \\ 1/2, & 1/4, & 1/3 \end{pmatrix} \\ & \downarrow \quad\quad \downarrow \quad\quad \downarrow \\ & 1 \quad\quad\; 1 \quad\quad\; 1 \end{array}$$

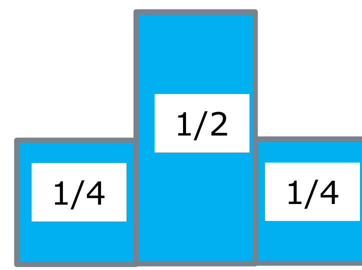
$x \in X$なる全てのxについて$\sum_{y \in Y} f_{yx} = 1$

$f(1) =$



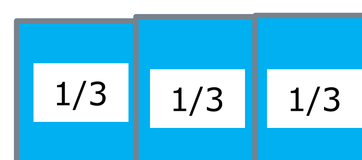
f_{11} f_{21} f_{31}

$f(2) =$



f_{12} f_{22} f_{32}

$f(3) =$



f_{13} f_{23} f_{33}

stochastic map $f: X \rightsquigarrow Y$ を行列で表す例 2

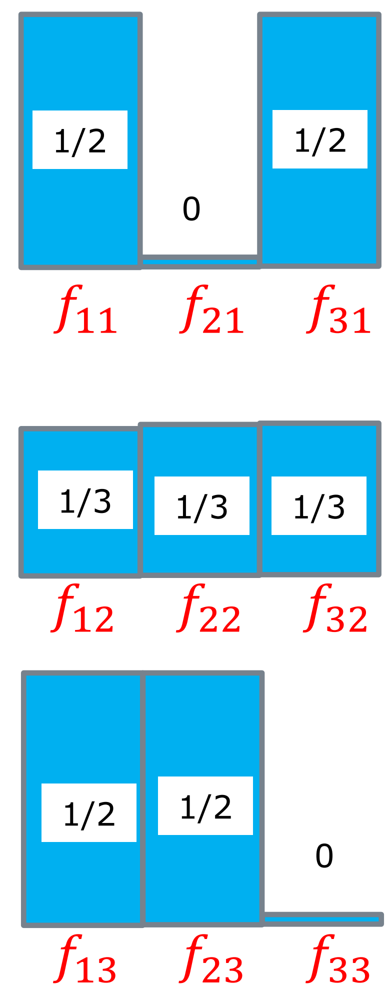
$$X \overset{f}{\rightsquigarrow} Y$$

$$
\begin{array}{c|ccc}
 & x=1, & x=2, & x=3 \\
\hline
y=1 & 1/2, & 1/3, & 1/2 \\
y=2 & 0, & 1/3, & 1/2 \\
y=3 & 1/2, & 1/3, & 0 \\
\hline
 & \downarrow & \downarrow & \downarrow \\
 & 1 & 1 & 1
\end{array}
$$

$x \in X$なる全てのxについて$\sum_{y \in Y} f_{yx} = 1$

f_{yx}

$f(1) =$, $f(2) =$, $f(3) =$

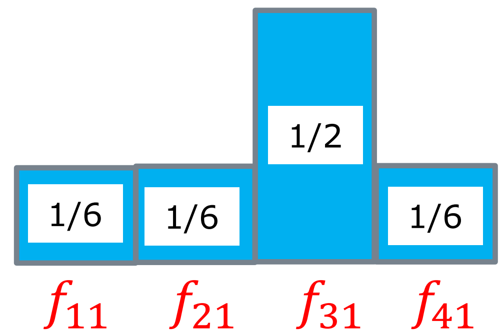


stochastic map $f : X \rightsquigarrow Y$ を行列で表す　例 3

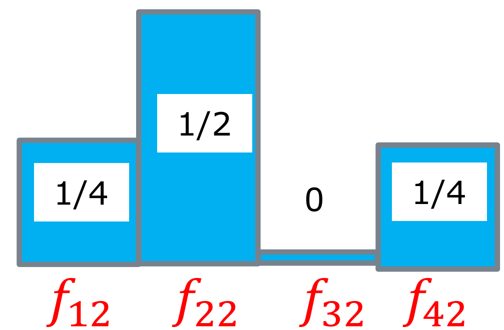
$$X \overset{f}{\rightsquigarrow} Y \qquad f_{yx}$$

$$\begin{array}{cc}
 & \begin{array}{ccc} x=1, & x=2, & x=3 \end{array} \\
\begin{array}{c} y=1 \\ y=2 \\ y=3 \\ y=4 \end{array} &
\begin{pmatrix} 1/6, & 1/4, & 1/4 \\ 1/6, & 1/2, & 1/4 \\ 1/2, & 0, & 1/4 \\ 1/6, & 1/4, & 1/4 \end{pmatrix} \\
 & \begin{array}{ccc} \downarrow & \downarrow & \downarrow \\ 1 & 1 & 1 \end{array}
\end{array}$$

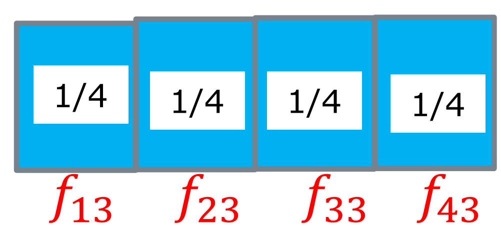
$f(1) =$



$f(2) =$



$f(3) =$



特別なstochastic map
$q : 1 \rightsquigarrow X$

一つしか要素を持たない集合を $\mathbf{1}$ で表す。

この時、stochastic map $q : \mathbf{1} \rightsquigarrow \boldsymbol{X}$ は、X上の確率分布 q を表す。以前の表記では、$(X, \boldsymbol{q})$、あるいは$\boldsymbol{q} = (q_1, q_2, \ldots q_m)$として$(X, (q_1, q_2, \ldots q_m))$と表していたものと同じである。

$$\begin{array}{c} \mathbf{1} \\ \Big\downarrow{\scriptstyle q} \\ X \end{array} \iff \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix} \iff \big(X, (q_1, q_2, \ldots q_m)\big) = (X, \boldsymbol{q})$$

on X

stochastic map $f : X \rightsquigarrow Y$ と stochastic map $g : Y \rightsquigarrow Z$ の合成

$f : X \rightsquigarrow Y$ と $g : Y \rightsquigarrow Z$ の合成 $g \circ f$ を、fを表す行列とgを表す行列の積で定義する。

$$(g \circ f)_{zx} = \sum_{y \in Y} g_{zy} f_{yx}$$

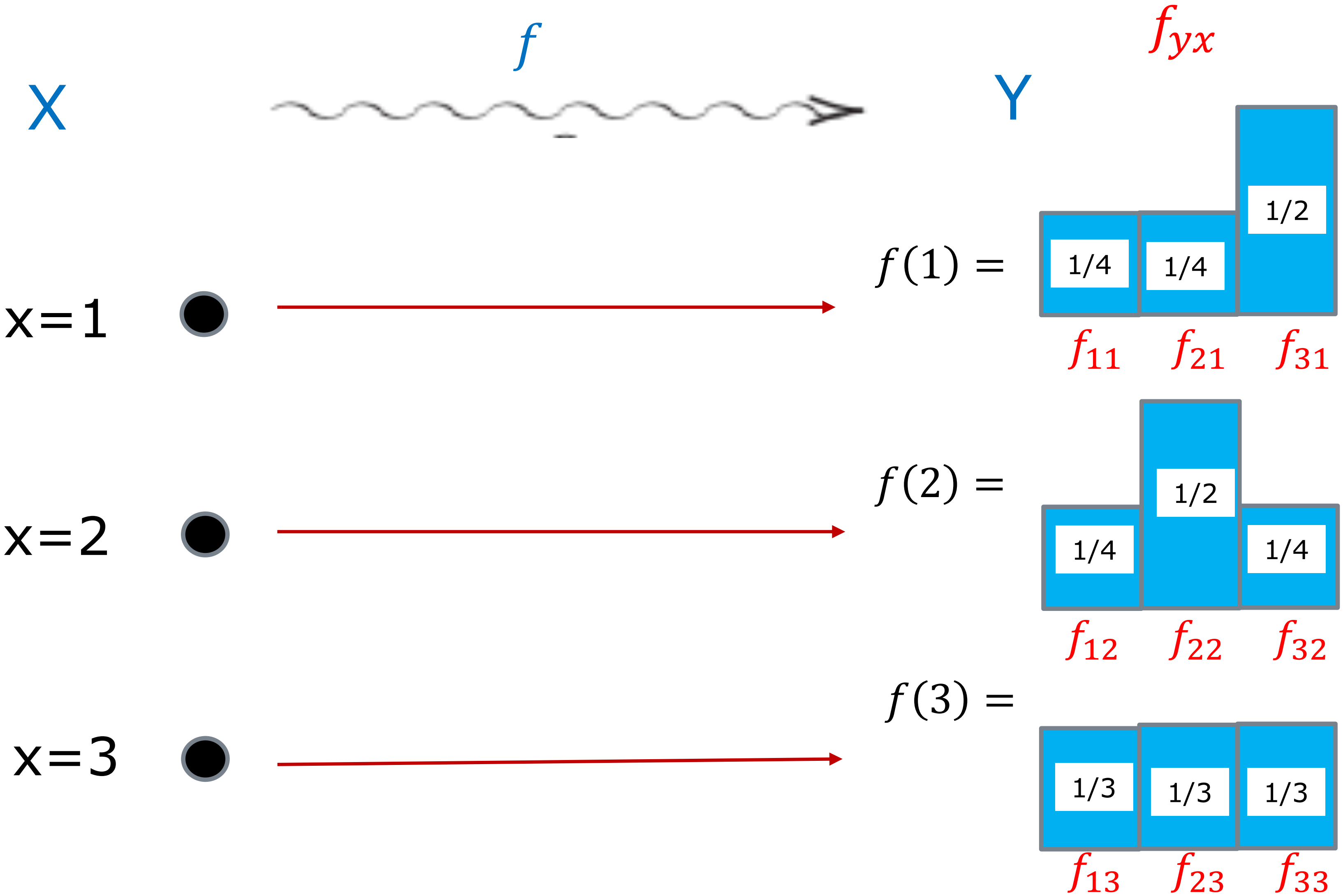
カテゴリー FinStoch

行列の積は結合的で、単位行列はstochastic なので、先に見た行列の積で射の合成が定義されるstochastic map は、カテゴリーの射としての要件を満たす。

有限集合$X, Y, Z, \ldots$をオブジェクトとし、stochastic map を射とするカテゴリーを**FinStoch**と呼ぶ。

stochastic mapと**FinProb**

stochastic map $f : X \rightsquigarrow Y$
Xの要素にYの確率分布に対応させる

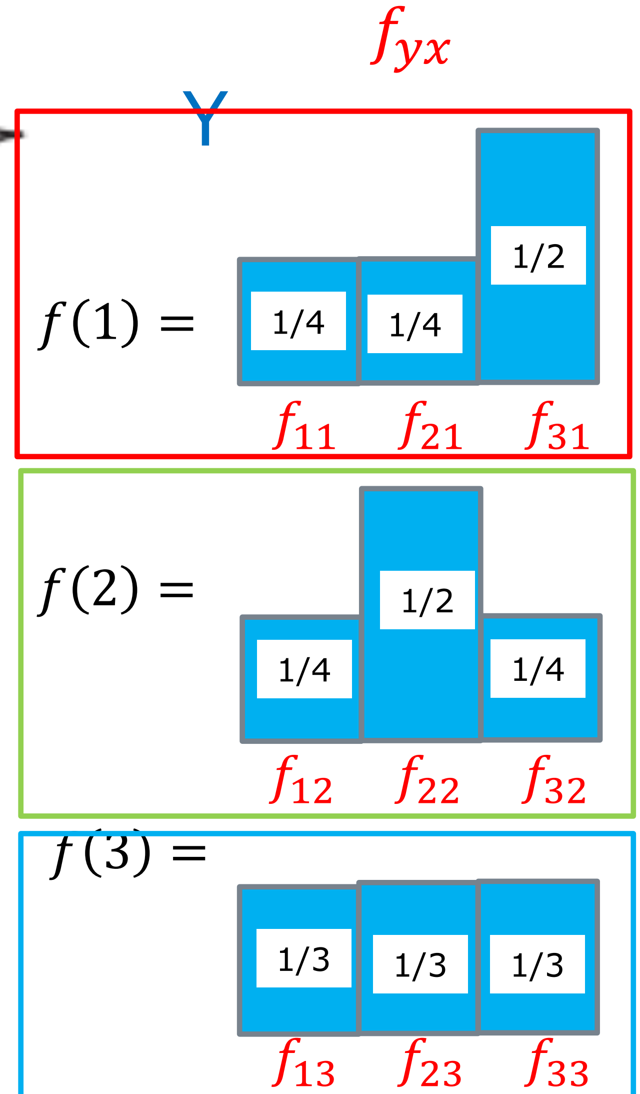


stochastic map $f : X \rightsquigarrow Y$ を行列で表す

$X \overset{f}{\rightsquigarrow} Y$

$$
\begin{array}{c|ccc}
 & x=1, & x=2, & x=3 \\
\hline
y=1 & 1/4, & 1/4, & 1/3 \\
y=2 & 1/4, & 1/2, & 1/3 \\
y=3 & 1/2, & 1/4, & 1/3 \\
\hline
\sum & \downarrow 1 & \downarrow 1 & \downarrow 1
\end{array}
$$

$x \in X$なる全てのxについて$\sum_{y \in Y} f_{yx} = 1$

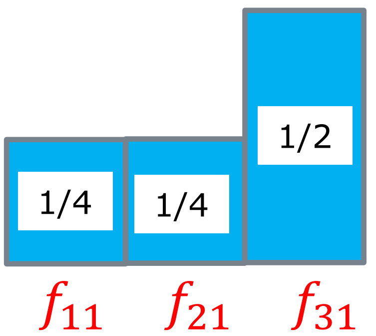


stochastic map $f : 1 \rightsquigarrow Y$ を
1は一つの要素しか持たない

$$1 \overset{f}{\rightsquigarrow} Y$$

f_{yx}

$f(1) =$



$$\begin{array}{cc} & x=1 \\ \begin{array}{c} y=1 \\ y=2 \\ y=3 \end{array} & \begin{pmatrix} 1/4 \\ 1/4 \\ 1/2 \end{pmatrix} \end{array}$$

↓

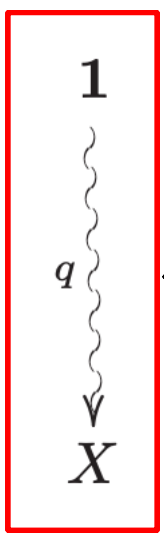
1

$x \in X$なる全てのxについて$\sum_{y \in Y} f_{yx} = 1$

特別なstochastic map
$q : 1 \rightsquigarrow X$

一つしか要素を持たない集合を $\mathbf{1}$ で表す。

この時、stochastic map $q : \mathbf{1} \rightsquigarrow \boldsymbol{X}$ は、X上の確率分布 q を表す。以前の表記では、$(X, \boldsymbol{q})$、あるいは$\boldsymbol{q} = (q_1, q_2, \ldots q_m)$として$(X, (q_1, q_2, \ldots q_m))$と表していたものと同じである。



$$\Longleftrightarrow \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix} \Longleftrightarrow \big(X, (q_1, q_2, \ldots q_m)\big) = (X, \boldsymbol{q})$$

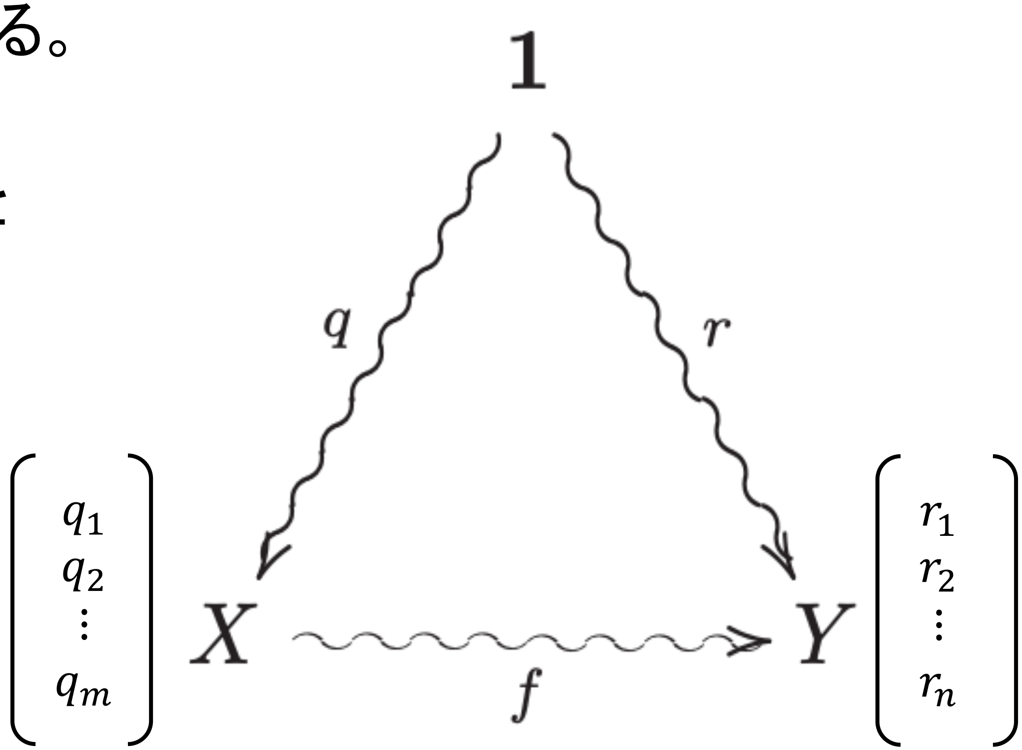
on X

stochastic mapの次の図式が可換だとする $r = f \circ q$

qはX上の確率分布**q**を与え、
rはY上の確率分布**r**を与える。

map fの行列をf_{yx}とすると
$r = f \circ q$より、

$$\boxed{r_y = \sum_{x \in X} f_{yx}\, q_x}$$



q on X　　**r** on Y

stochastic mapで **FinProb**を再構成する

fが関数の時、fの行列は、

$$f_{yx} = \delta_{yf(x)}$$

δはクロネッカーのデルタ

で表されるから、

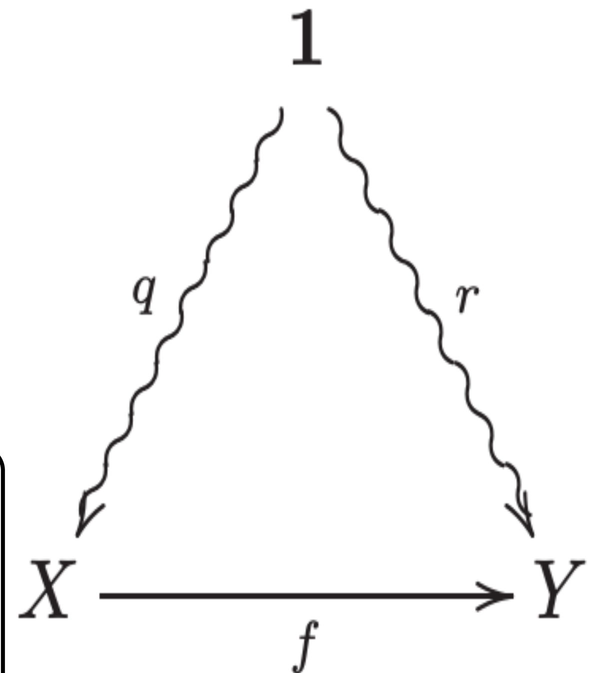
$$r_y = \sum_{x \in X} f_{yx}\, q_x$$

$$= \sum_{x \in X} \delta_{yf(x)}\, q_x$$

$$= \sum_{x : f(x) = y} q_x$$

すなわち、fは測度を保存する関数である。

これは、**FinProb**の図式である



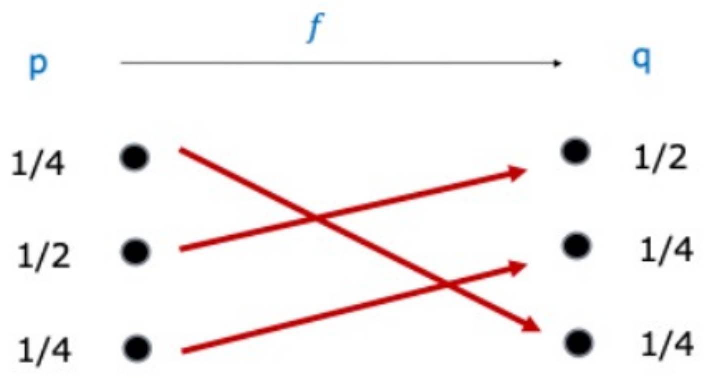
$$\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix} \qquad \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

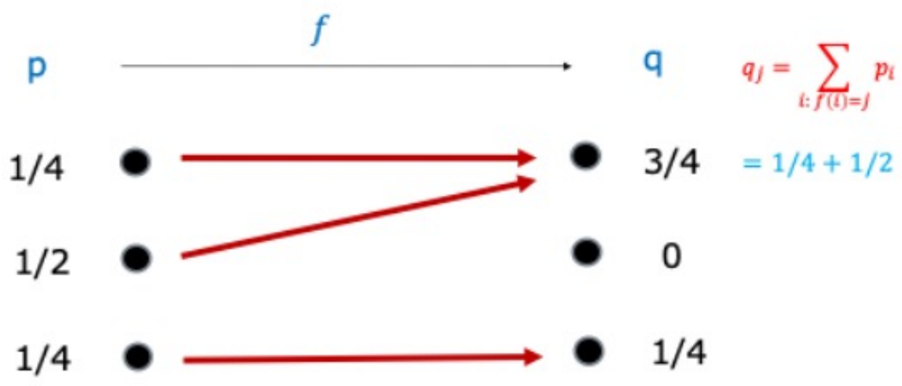
fは測度を保存する関数

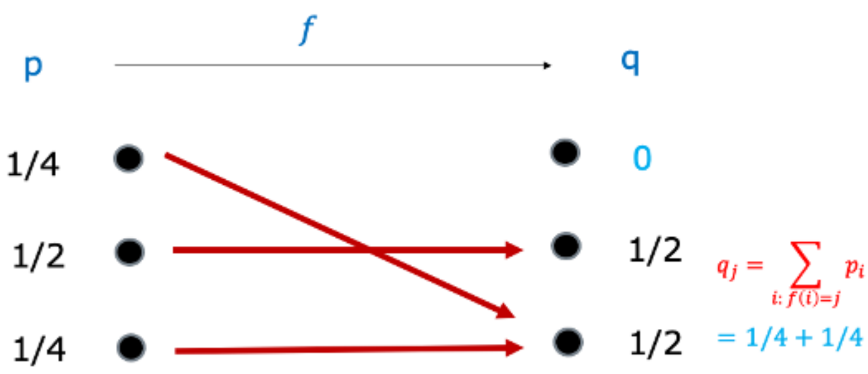
X上の確率分布q

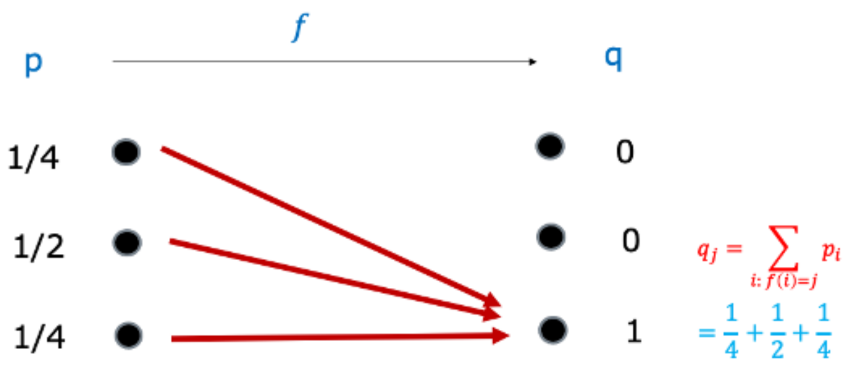
Y上の確率分布r

FinProbと
測度を保存する関数 f の例

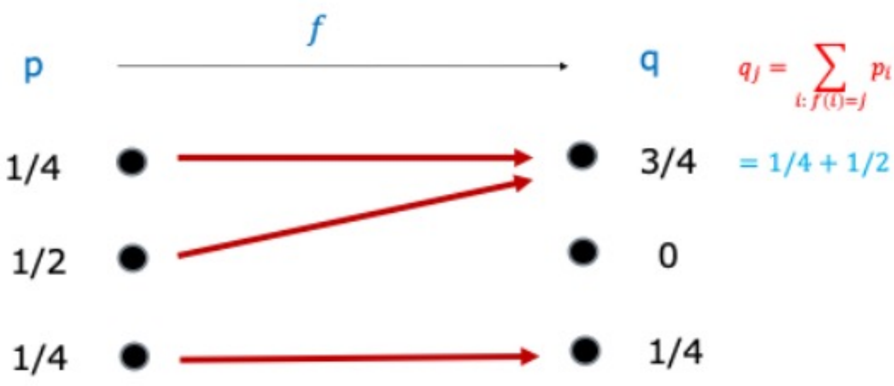








関数 $f: X \to Y$を行列で表す

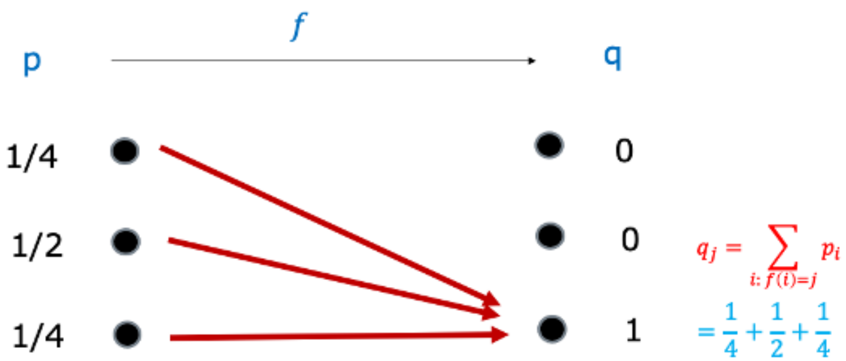


$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix} = \begin{pmatrix} 3/4 \\ 0 \\ 1/4 \end{pmatrix}$$

$$f_{yx} = \delta_{y f(x)}$$
$$f(y, x) = \delta(y, f(x))$$
$$f(1,1) = \delta(1, f(1)) = \delta(1,1) = 1$$
$$f(1,2) = \delta(1, f(2)) = \delta(1,1) = 1$$
$$f(1,3) = \delta(1, f(3)) = \delta(1,3) = 0$$
$$f(2,1) = \delta(2, f(1)) = \delta(2,1) = 0$$
$$f(2,2) = \delta(2, f(2)) = \delta(2,1) = 0$$
$$f(2,3) = \delta(2, f(3)) = \delta(2,3) = 0$$
$$f(3,1) = \delta(3, f(1)) = \delta(3,1) = 0$$
$$f(3,2) = \delta(3, f(2)) = \delta(3,1) = 0$$
$$f(3,3) = \delta(3, f(3)) = \delta(3,3) = 1$$

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

関数 $f: X \to Y$を行列で表す



$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{bmatrix} 1/4 \\ 1/2 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$f_{yx} = \delta_{y f(x)}$$
$$f(y,x) = \delta(y, f(x))$$
$$f(1,1) = \delta\big(1, f(1)\big) = \delta(1,3) = 0$$
$$f(1,2) = \delta\big(1, f(2)\big) = \delta(1,3) = 0$$
$$f(1,3) = \delta\big(1, f(3)\big) = \delta(1,3) = 0$$
$$f(2,1) = \delta\big(2, f(1)\big) = \delta(2,3) = 0$$
$$f(2,2) = \delta\big(2, f(2)\big) = \delta(2,3) = 0$$
$$f(2,3) = \delta\big(2, f(3)\big) = \delta(2,3) = 0$$
$$f(3,1) = \delta\big(3, f(1)\big) = \delta(3,3) = 1$$
$$f(3,2) = \delta\big(3, f(2)\big) = \delta(3,3) = 1$$
$$f(3,3) = \delta\big(3, f(3)\big) = \delta(3,3) = 1$$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

FinProb

$$\begin{tikzcd}
 & & \mathbf{1} \arrow[dl, squiggly, "q"'] \arrow[dr, squiggly, "r"] & & \\
\begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_m \end{pmatrix} & X \arrow[rr, "f"'] & & Y & \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}
\end{tikzcd}$$

fは測度を保存する関数

X上の確率分布q

Y上の確率分布r

FinStatと「仮説」

カテゴリー **FinStat**

相対エントロピーのカテゴリー論的特徴づけに利用されるカテゴリー **Finstat** は、すこし複雑な形をしている。いよいよ、それをみていこう。

カテゴリー論らしく、**Finstat** のオブジェクトと射を、図式でみていこう。

カテゴリー FinStat の
オブジェクトと射

Finstatのオブジェクト

$$
\begin{tikzcd}
\mathbf{1} \arrow[d, squiggly, "q"'] \\
X
\end{tikzcd}
$$

Finstatの射

$$
\begin{tikzcd}
& \mathbf{1} \arrow[dl, squiggly, "q"'] \arrow[dr, squiggly, "r"] & \\
X \arrow[rr, bend right, "f"'] & & Y \arrow[ll, squiggly, bend right, "s"']
\end{tikzcd}
$$

$$
\begin{aligned}
f \circ q &= r \\
f \circ s &= 1_Y
\end{aligned}
$$

カテゴリー FinStat のオブジェクト

Finstatのオブジェクト

$$\begin{tikzcd} \mathbf{1} \arrow[d, squiggly, "q"'] \\ X \end{tikzcd}$$

Finstatのオブジェクトは、有限集合上の確率測度の空間である。

カテゴリー FinStat の射

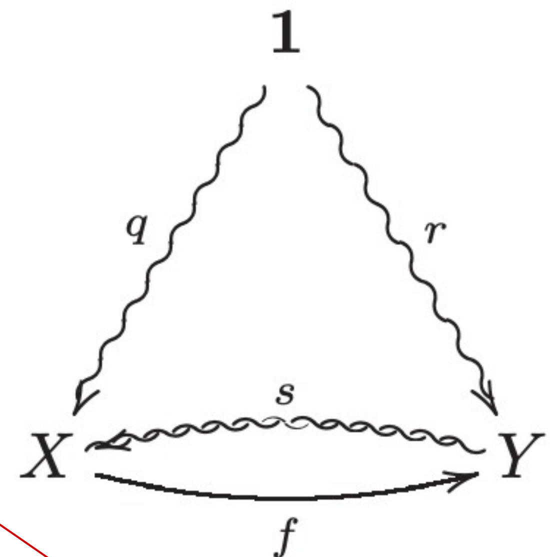
この図式は、可換である必要はない。

ただし、二つの条件は必須である。

最初の条件は、$f : X \to Y$ が、測度を保存する関数であることを主張している。これは、前回見た。

第二の条件 は、$f \circ s$ がYの上の同一性の射であることを主張している。これをsは"section"であるともいう。

Finstatの射



$$
\begin{aligned}
f \circ q &= r \\
f \circ s &= 1_Y
\end{aligned}
$$

カテゴリー FinStat の解釈

Xは、あるシステムの「状態」と考えることができる。一方のYは、別のシステム -- 「観測機器」の可能な状態の集合である。

関数 f は、「観測過程」である。我々は f を使ってシステムを観測する。もしシステムが状態$x \in X$にあれば、観測機器の状態は$f(x)$になる。

確率分布qは、システムが与えられた任意の状態にある確率を与える。

一方、確率分布rは、観測がなされた後に観測機器が取る与えられた任意の状態の確率を与える。

Finstat

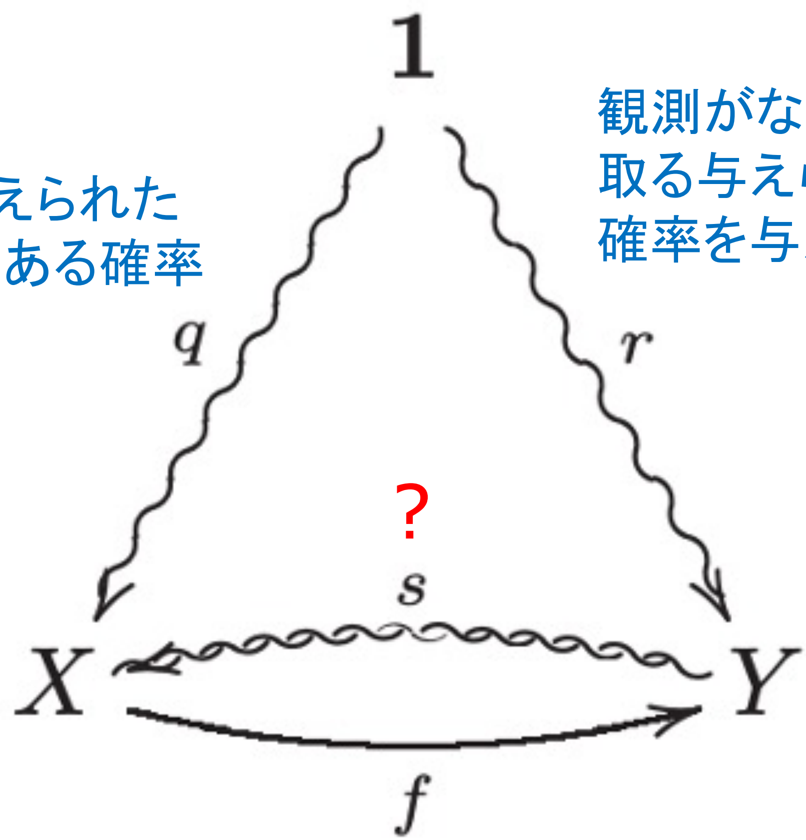
$$
\begin{tikzcd}
 & \mathbf{1} \arrow[dl, squiggly, "q"'] \arrow[dr, squiggly, "r"] & \\
X \arrow[rr, bend right=10, "f"'] & & Y \arrow[ll, squiggly, bend right=10, "s"']
\end{tikzcd}
$$

$$
\begin{aligned}
f \circ q &= r \\
f \circ s &= 1_Y
\end{aligned}
$$

FinStat

システムが与えられた
任意の状態にある確率
を与える

観測がなされた後に観測機器が
取る与えられた任意の状態の
確率を与える



あるシステムの
取る状態

観測機器の
取る状態

観測過程

s を「仮説」として解釈する

我々は、stochastic map s を、観測機器で与えられた状態のもとでのシステムの状態についての「仮説」と考える。

システムを観察して、観測機器が $y \in Y$ の状態をとった時、この「仮説」は、システムが確率 s_{yx} で x の状態にあったことを主張する。

式 $f \circ s = 1_Y$ は、もしある観測で観測機器が$y \in Y$の状態をとった時、この「仮説」は、この状態 $y \in Y$を実際に与える観測で、観測されたシステムの状態にだけゼロでない確率 $x \in X$を割り当てることを意味する。

$f \circ s = 1_Y$ となる条件

「仮説」にとって重要な条件 $f \circ s = 1_Y$が成り立つのは、どういう場合なのかを考えてみよう。

次のことが言える。

$f\colon X \to Y$ が有限集合 X, Y の間の関数で、
$s\colon Y \rightsquigarrow X$ がそれらの間のstochastic map だとする。この時、

$f \circ s = 1_Y$ が成り立つのは、全ての$y \in Y$に対して、$f(x) = y$ でなければ $s_{xy} = 0$である場合であり、かつ、その場合に限る。

$f \circ s = 1_Y$ となる条件をすこし詳しく見ておこう

任意の$y, y' \in Y$について、$f \circ s = 1_Y$が成り立っているとしよう。

$$(f \circ s)_y = \sum_{x \in X} f_{yx} s_{xy} = \sum_{x \in X} \delta_{y f(x)} s_{xy} = 1$$

あるxについて　$f(x) = y$とすれば、$\delta_{y f(x)} = 1$

この時、$(f \circ s)_y = \sum_{x \in X} \delta_{y f(x)} s_{xy} = \sum_{x \in X} s_{xy} = 1$

fは関数なので、$y \neq y'$なるy'について　$f(x) \neq y'$

この時、$\delta_{y' f(x)} = 0$ だから、$(f \circ s)_{y'} = \sum_{x \in X} \delta_{y' f(x)} s_{xy} = 0$

s がstochastic map ならば、$s_{xy} \geq 0$ であるので、

これは、全ての$y \in Y$に対して、$f(x) = y$ でなければ $s_{xy} = 0$ を意味する。

$f \circ s = 1_Y$ となる条件を すこし詳しく見ておこう

逆に、$s: Y \rightsquigarrow X$ がstochastic map で、全ての$y \in Y$に対して、$f(x) = y$ でなければ $s_{xy} = 0$ だとする。

この時、$y \in Y$なる全てのyについて、次の式が成り立つ。

$$1 = \sum_{x \in X} s_{xy} = \sum_{x: f(x) = y} s_{xy} = \sum_{x \in X} \delta_{y\, f(x)} s_{xy}$$

一方、$y \neq y'$ならば、$f(x) \neq y'$

$$0 = \sum_{x \in X} \delta_{y' f(x)} s_{xy}$$

$$\therefore \sum_{x \in X} \delta_{y' f(x)} s_{xy} = \sum_{x \in X} f_{y' x}\, s_{xy} = \delta_{y' y}$$

これは、$f \circ s = 1_Y$ を意味する。

二つの射の合成

二つの射の合成を次のように定義する。

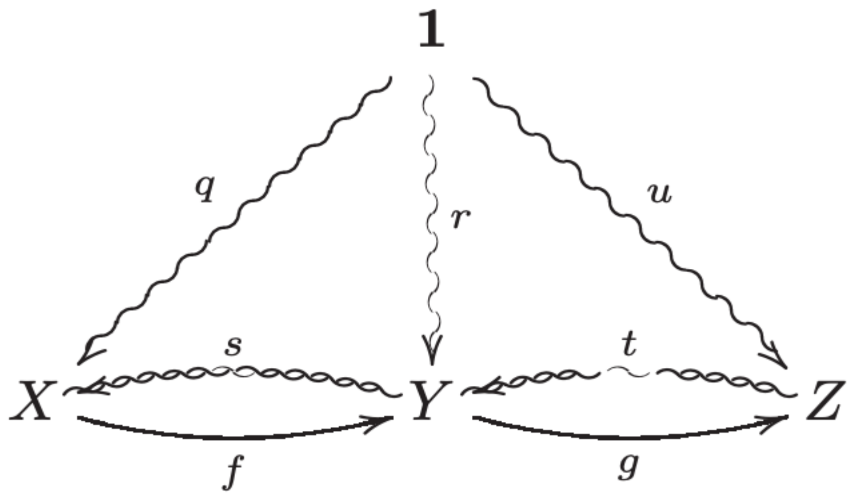
$$
\begin{tikzcd}
 & & \mathbf{1} \arrow[dll, squiggly, "q"'] \arrow[d, squiggly, "r"] \arrow[drr, squiggly, "u"] & & \\
X \arrow[rr, bend right, "f"'] & & Y \arrow[ll, squiggly, "s"'] \arrow[rr, bend right, "g"'] & & Z \arrow[ll, squiggly, "t"']
\end{tikzcd}
$$

$$
\begin{aligned}
f \circ q &= r \qquad & g \circ r &= u \\
f \circ s &= 1_Y \qquad & g \circ t &= 1_Z
\end{aligned}
$$

測度を保存する関数 $g \circ f : X \to Z$と、
逆方向のstochastic map $s \circ t : Z \rightsquigarrow X$ をうる。

二つの射の合成

二つの射の合成を次のように定義する。



$$
\begin{aligned}
f \circ q &= r \qquad & g \circ r &= u \\
f \circ s &= 1_Y \qquad & g \circ t &= 1_Z
\end{aligned}
$$

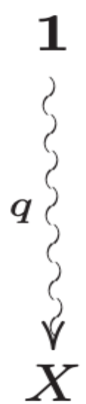
この時、次の条件は満たされている。

$$g \circ f \circ q = u$$

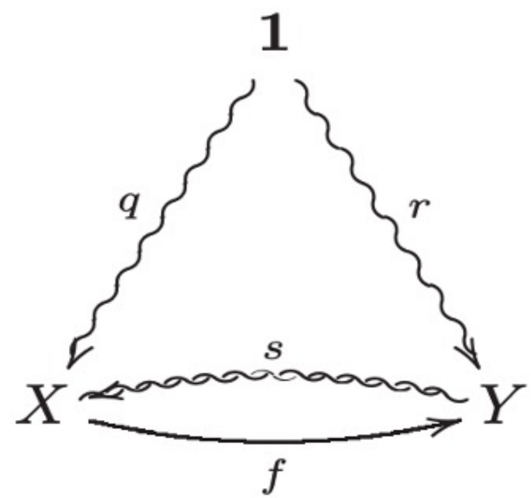
$$g \circ f \circ s \circ t = 1_Z$$

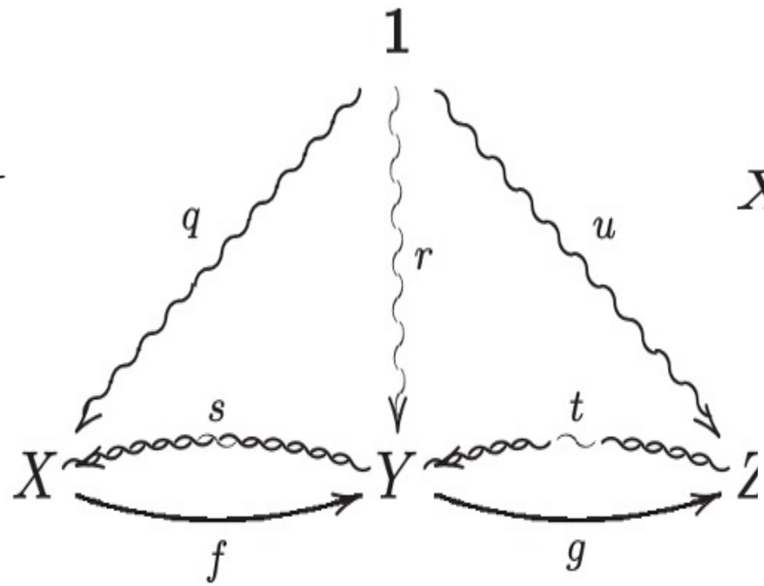
カテゴリー **FinStat** の
オブジェクトと射と射の合成

Finstatのオブジェクト



Finstatの射





Finstatの射の合成

FinStatと「仮説」

カテゴリー **FinStat**

相対エントロピーのカテゴリー論的特徴づけに利用されるカテゴリー **Finstat** は、すこし複雑な形をしている。いよいよ、それをみていこう。

カテゴリー論らしく、**Finstat** のオブジェクトと射を、図式でみていこう。

カテゴリー FinStat の
オブジェクトと射

Finstatのオブジェクト

$$
\begin{tikzcd}
\mathbf{1} \arrow[d, rightsquigarrow, "q"'] \\
X
\end{tikzcd}
$$

Finstatの射

$$
\begin{tikzcd}
 & \mathbf{1} \arrow[dl, rightsquigarrow, "q"'] \arrow[dr, rightsquigarrow, "r"] & \\
X \arrow[rr, bend right, "f"'] & & Y \arrow[ll, rightsquigarrow, bend right, "s"']
\end{tikzcd}
$$

$$
\begin{aligned}
f \circ q &= r \\
f \circ s &= 1_Y
\end{aligned}
$$

カテゴリー FinStat のオブジェクト

Finstatのオブジェクト

$$\begin{tikzcd} \mathbf{1} \arrow[d, squiggly, "q"'] \\ X \end{tikzcd}$$

Finstatのオブジェクトは、有限集合上の確率測度の空間である。

カテゴリー FinStat の射

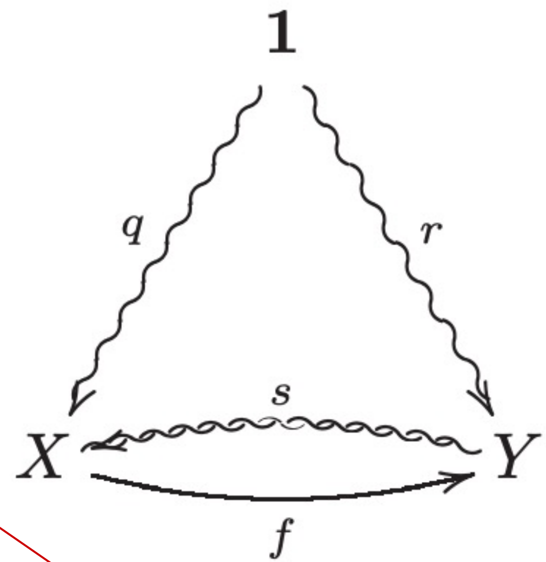
Finstatの射

この図式は、可換である必要はない。

ただし、二つの条件は必須である。

最初の条件は、$f : X \to Y$ が、測度を保存する関数であることを主張している。これは、前回見た。

第二の条件 は、$f \circ s$ がYの上の同一性の射であることを主張している。これをsは"section"であるともいう。



$$\begin{aligned} f \circ q &= r \\ f \circ s &= 1_Y \end{aligned}$$

カテゴリー FinStat の解釈

Xは、あるシステムの「状態」と考えることができる。一方のYは、別のシステム --「観測機器」の可能な状態の集合である。

関数 f は、「観測過程」である。我々は f を使ってシステムを観測する。もしシステムが状態$x \in X$にあれば、観測機器の状態は$f(x)$になる。

確率分布qは、システムが与えられた任意の状態にある確率を与える。

一方、確率分布rは、観測がなされた後に観測機器が取る与えられた任意の状態の確率を与える。

Finstat

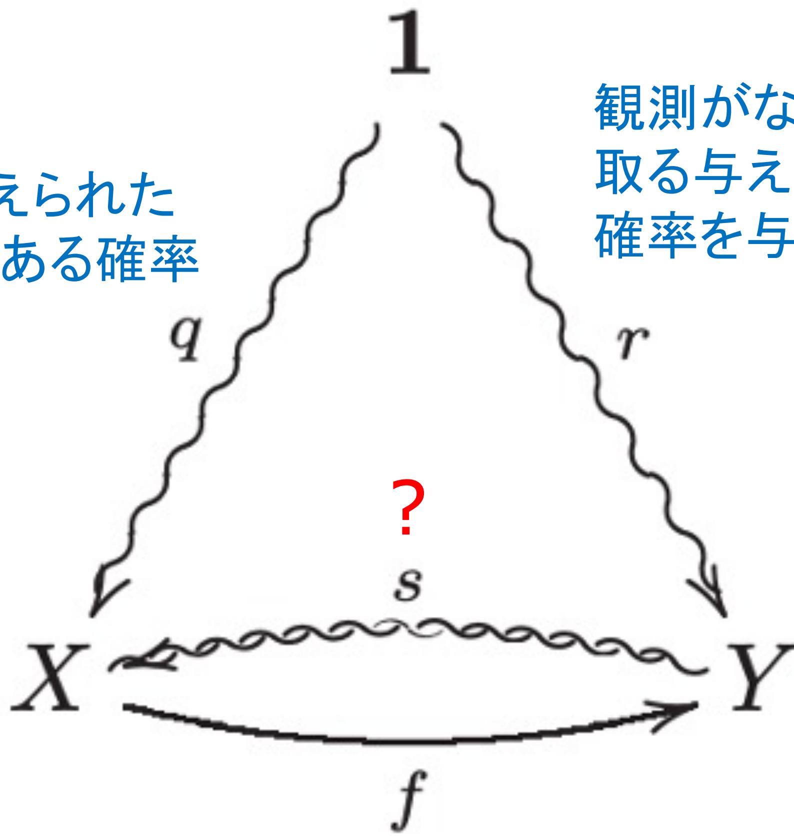
$$\begin{tikzcd}
 & \mathbf{1} \arrow[dl, rightsquigarrow, "q"'] \arrow[dr, rightsquigarrow, "r"] & \\
X \arrow[rr, bend right, "f"'] & & Y \arrow[ll, rightsquigarrow, "s"']
\end{tikzcd}$$

$$\begin{aligned} f \circ q &= r \\ f \circ s &= 1_Y \end{aligned}$$

FinStat

システムが与えられた
任意の状態にある確率
を与える

観測がなされた後に観測機器が
取る与えられた任意の状態の
確率を与える



あるシステムの
取る状態

観測機器の
取る状態

観測過程

sを「仮説」として解釈する

我々は、stochastic map sを、観測機器で与えられた状態のもとでのシステムの状態についての「仮説」と考える。

システムを観察して、観測機器が $y \in Y$の状態をとった時、この「仮説」は、システムが確率s_{yx}でxの状態にあったことを主張する。

式 $f \circ s = 1_Y$ は、もしある観測で観測機器が$y \in Y$の状態をとった時、この「仮説」は、この状態 $y \in Y$を実際に与える観測で、観測されたシステムの状態にだけゼロでない確率 $x \in X$を割り当てることを意味する。

$f \circ s = 1_Y$ となる条件

「仮説」にとって重要な条件 $f \circ s = 1_Y$ が成り立つのは、どういう場合なのかを考えてみよう。

次のことが言える。

$f: X \to Y$ が有限集合 X, Y の間の関数で、
$s: Y \rightsquigarrow X$ がそれらの間のstochastic map だとする。この時、

$f \circ s = 1_Y$ が成り立つのは、全ての$y \in Y$に対して、$f(x) = y$ でなければ $s_{xy} = 0$である場合であり、かつ、その場合に限る。

$f \circ s = 1_Y$ となる条件を すこし詳しく見ておこう

任意の$y, y' \in Y$について、$f \circ s = 1_Y$が成り立っているとしよう。

$$(f \circ s)_y = \sum_{x \in X} f_{yx} s_{xy} = \sum_{x \in X} \delta_{y f(x)} s_{xy} = 1$$

あるxについて　$f(x) = y$とすれば、$\delta_{y f(x)} = 1$

この時、$(f \circ s)_y = \sum_{x \in X} \delta_{y f(x)} s_{xy} = \sum_{x \in X} s_{xy} = 1$

fは関数なので、$y \neq y'$なるy'について　$f(x) \neq y'$

この時、$\delta_{y' f(x)} = 0$ だから、$(f \circ s)_{y'} = \sum_{x \in X} \delta_{y' f(x)} s_{xy} = 0$

s がstochastic map ならば、$s_{xy} \geq 0$ であるので、

これは、全ての$y \in Y$に対して、$f(x) = y$ でなければ $s_{xy} = 0$ を意味する。

$f \circ s = 1_Y$ となる条件をすこし詳しく見ておこう

逆に、$s: Y \rightsquigarrow X$ がstochastic map で、全ての$y \in Y$に対して、$f(x) = y$ でなければ $s_{xy} = 0$ だとする。

この時、$y \in Y$なる全てのyについて、次の式が成り立つ。

$$1 = \sum_{x \in X} s_{xy} = \sum_{x : f(x) = y} s_{xy} = \sum_{x \in X} \delta_{y\, f(x)} s_{xy}$$

一方、$y \neq y'$ならば、$f(x) \neq y'$

$$0 = \sum_{x \in X} \delta_{y' f(x)} s_{xy}$$

$$\therefore \sum_{x \in X} \delta_{y' f(x)} s_{xy} = \sum_{x \in X} f_{y' x}\, s_{xy} = \delta_{y' y}$$

これは、$f \circ s = 1_Y$ を意味する。

二つの射の合成

二つの射の合成を次のように定義する。

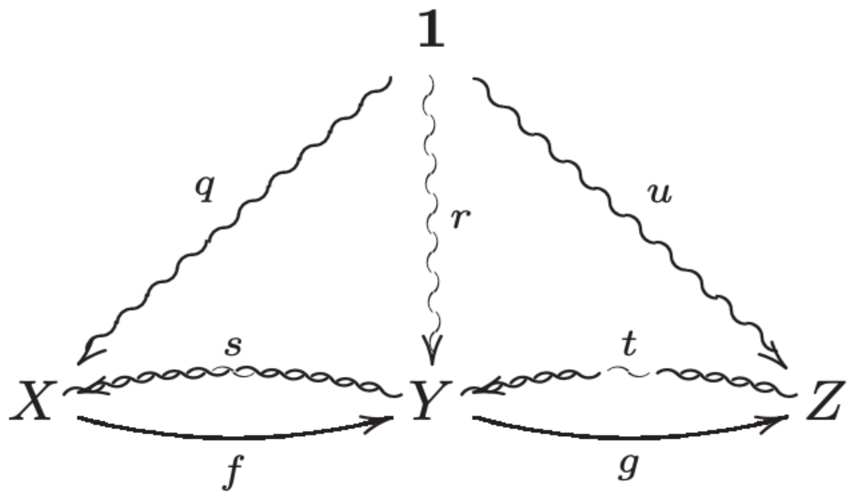
$$
\begin{tikzcd}
& & \mathbf{1} \arrow[dll, squiggly, "q"'] \arrow[d, squiggly, "r"] \arrow[drr, squiggly, "u"] & & \\
X \arrow[rr, bend right, "f"'] & & Y \arrow[ll, squiggly, "s"'] \arrow[rr, bend right, "g"'] & & Z \arrow[ll, squiggly, "t"']
\end{tikzcd}
$$

$$
\begin{aligned}
f \circ q &= r & \qquad g \circ r &= u \\
f \circ s &= 1_Y & \qquad g \circ t &= 1_Z
\end{aligned}
$$

測度を保存する関数 $g \circ f : X \to Z$と、
逆方向のstochastic map $s \circ t : Z \rightsquigarrow X$ をうる。

二つの射の合成

二つの射の合成を次のように定義する。



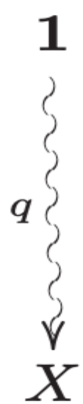
$$
\begin{aligned}
f \circ q &= r & \qquad g \circ r &= u \\
f \circ s &= 1_Y & \qquad g \circ t &= 1_Z
\end{aligned}
$$

この時、次の条件は満たされている。

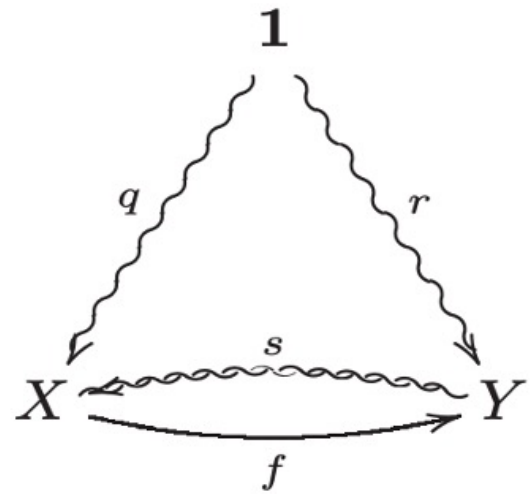
$$g \circ f \circ q = u$$
$$g \circ f \circ s \circ t = 1_Z$$

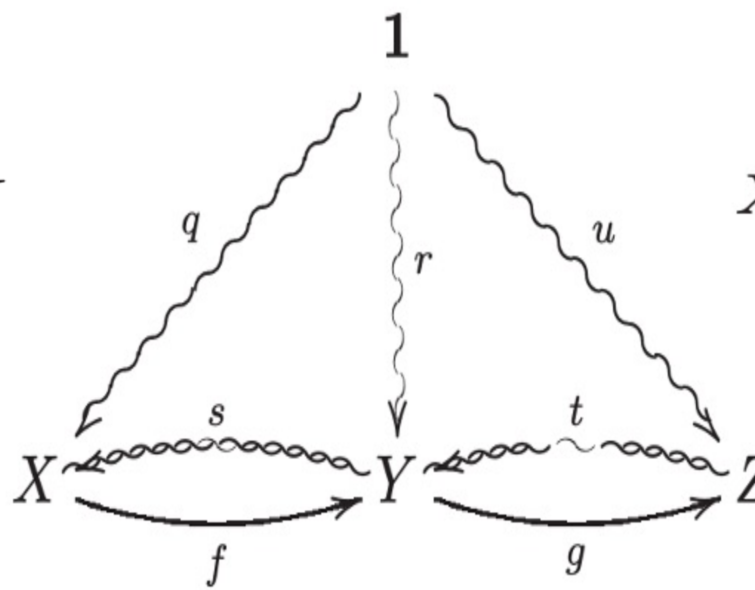
カテゴリー FinStat の
オブジェクトと射と射の合成

Finstatのオブジェクト



Finstatの射





Finstatの射の合成

第二部

エントロピー概念の拡大

第二部
エントロピー概念の拡大

1. Tsallis エントロピー
2. カテゴリーとして実数を捉える
3. Tsallis エントロピー = q-$\boldsymbol{logarithmic}$エントロピー
4. RényiエントロピーとHill数

Tsallis エントロピー

拡大されたエントロピー概念

$$\boldsymbol{Shannon, Boltzmann, Gibbs}$$

$$H = -\sum_i p_i \log p_i$$

$$\boldsymbol{Tsallis\ (q - logarithmic)}$$

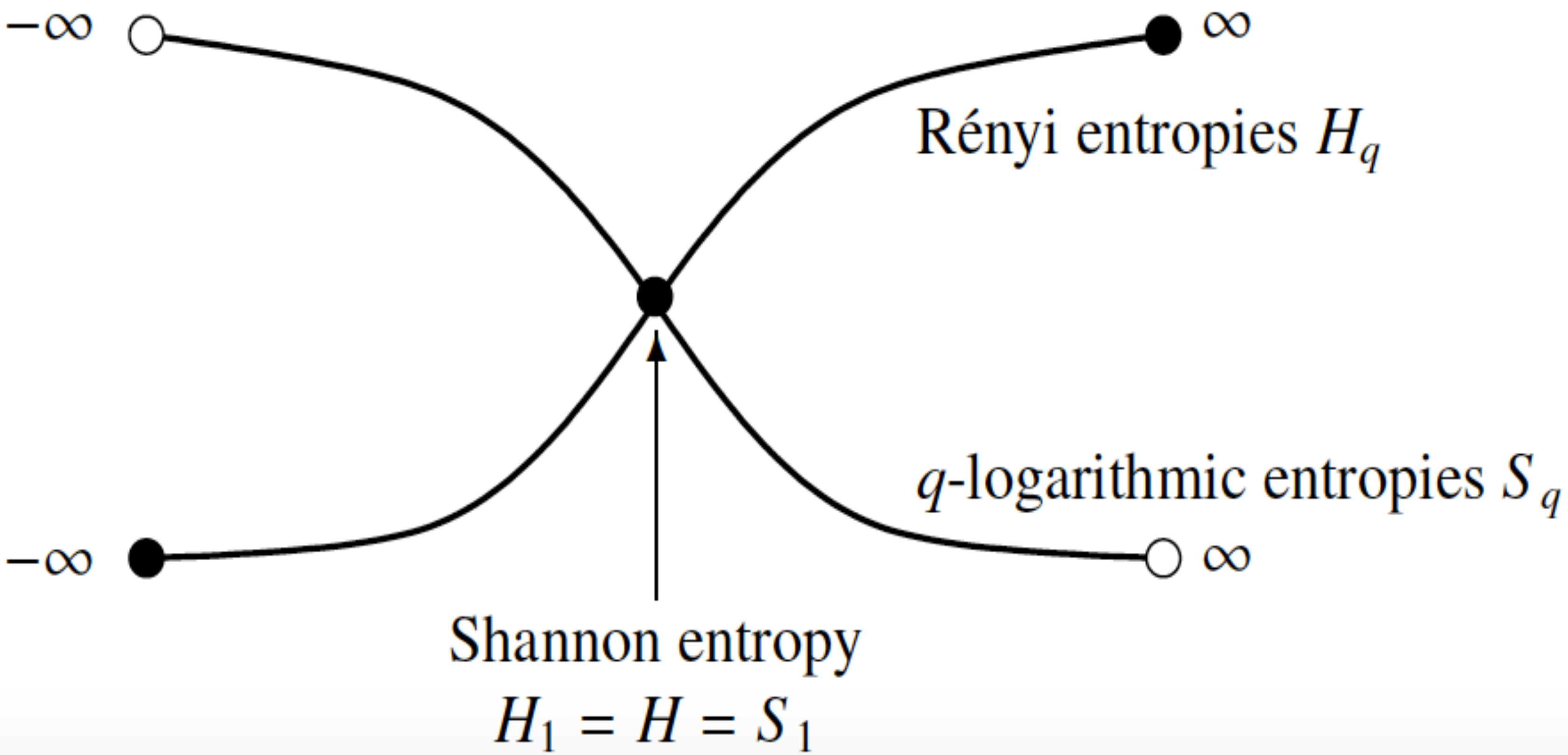
$$S_q = \frac{k}{q-1}\left(1 - \sum_i p_i^q\right)$$

$$\boldsymbol{Rényi}$$

$$H_q = \frac{1}{1-q}\log\left(\sum_i p_i^q\right)$$

$q \to 1$ の時

$H_1 = H = S_1$



Tsallis エントロピーの性質

シャノンエントロピー: extensive

二つの独立した(共通部分を)もたない領域 A,Bのエントロピー H(A,B)はAのエントロピーとBのエントロピーの和である。加法的

$$H(A,B) = H(A) + H(B)$$

(こうした性質をもつ量を、extensiveな量という.日本語だと「示量性」という。そうでない量の性質を intensive 「示強性」という)

Tsallis エントロピー: non-extensive

$$S_q(A,b) = S_q(A) + S_q(B) + (1-q)S_q(A)S_q(B)$$

Tsallis エントロピーは加法的でない。

なぜ、エントロピーの拡張が必要だったのか？ さまざまなアノマリーの存在と、その解決

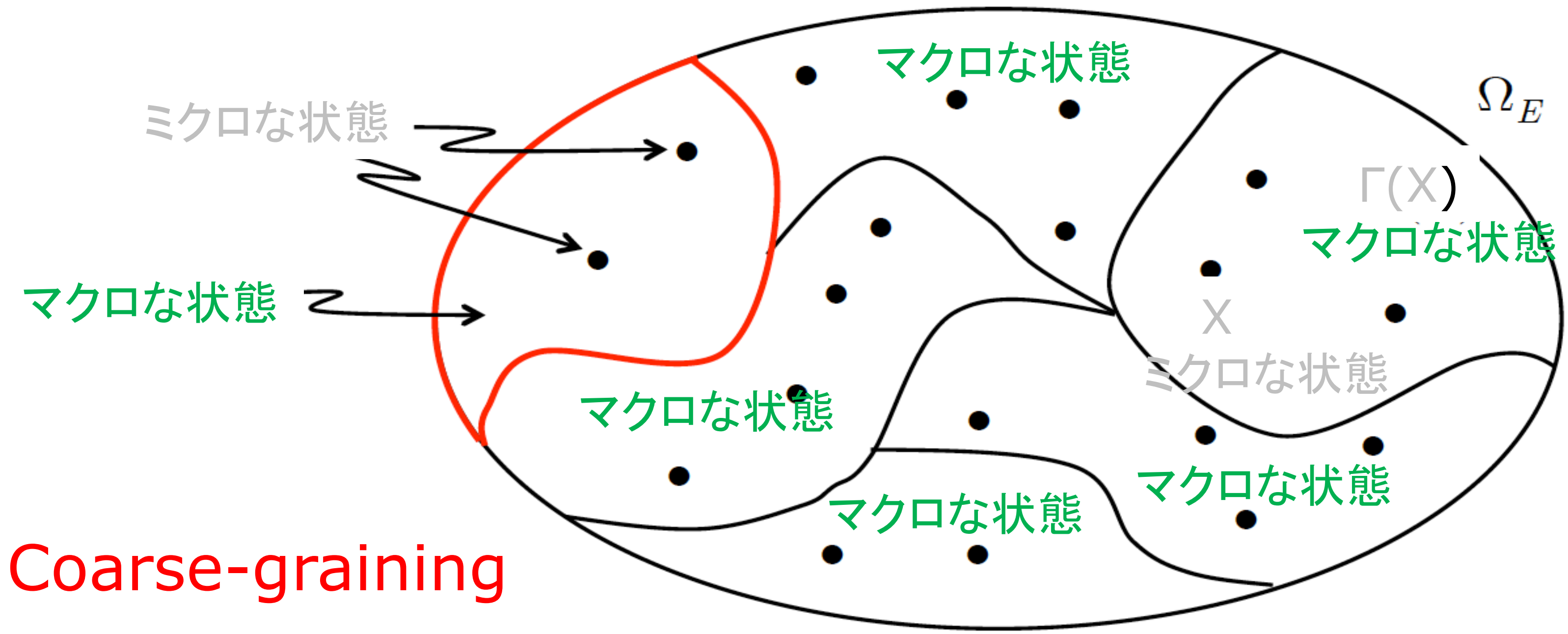
- The distribution characterizing the motion of cold atoms in dissipative optical lattices predicted in 2003[7] and observed in 2006.[8]
- The fluctuations of the magnetic field in the solar wind enabled the calculation of the q-triplet (or Tsallis triplet).[9]
- The velocity distributions in a driven dissipative dusty plasma.[10]
- Spin glass relaxation.[11]
- Trapped ion interacting with a classical buffer gas.[12]
- High energy collisional experiments at LHC/CERN (CMS, ATLAS and ALICE detectors)[13][14] and RHIC/Brookhaven (STAR and PHENIX detectors).[15]
- ….

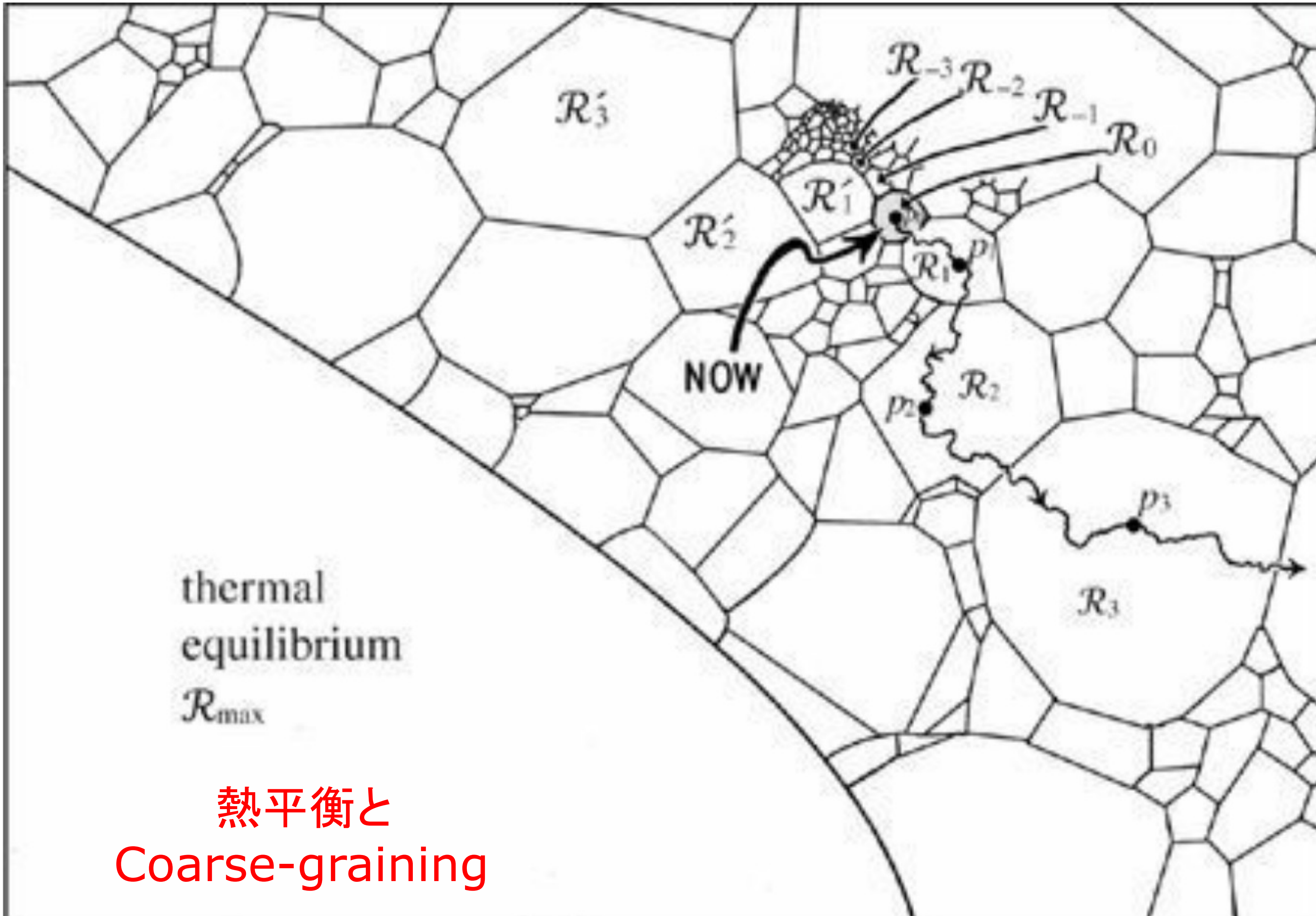
https://en.wikipedia.org/wiki/Tsallis_entropy

$\boldsymbol{Boltzmann, Gibbs}$のエントロピーの導出：ミクロな状態からマクロな状態へ

- 一つ一つのミクロな状態Xは、正確に、一つのマクロの状態Γ(X)に対応している。
- 多くの異なるミクロな状態が、一つのマクロな状態に対応づけられる。

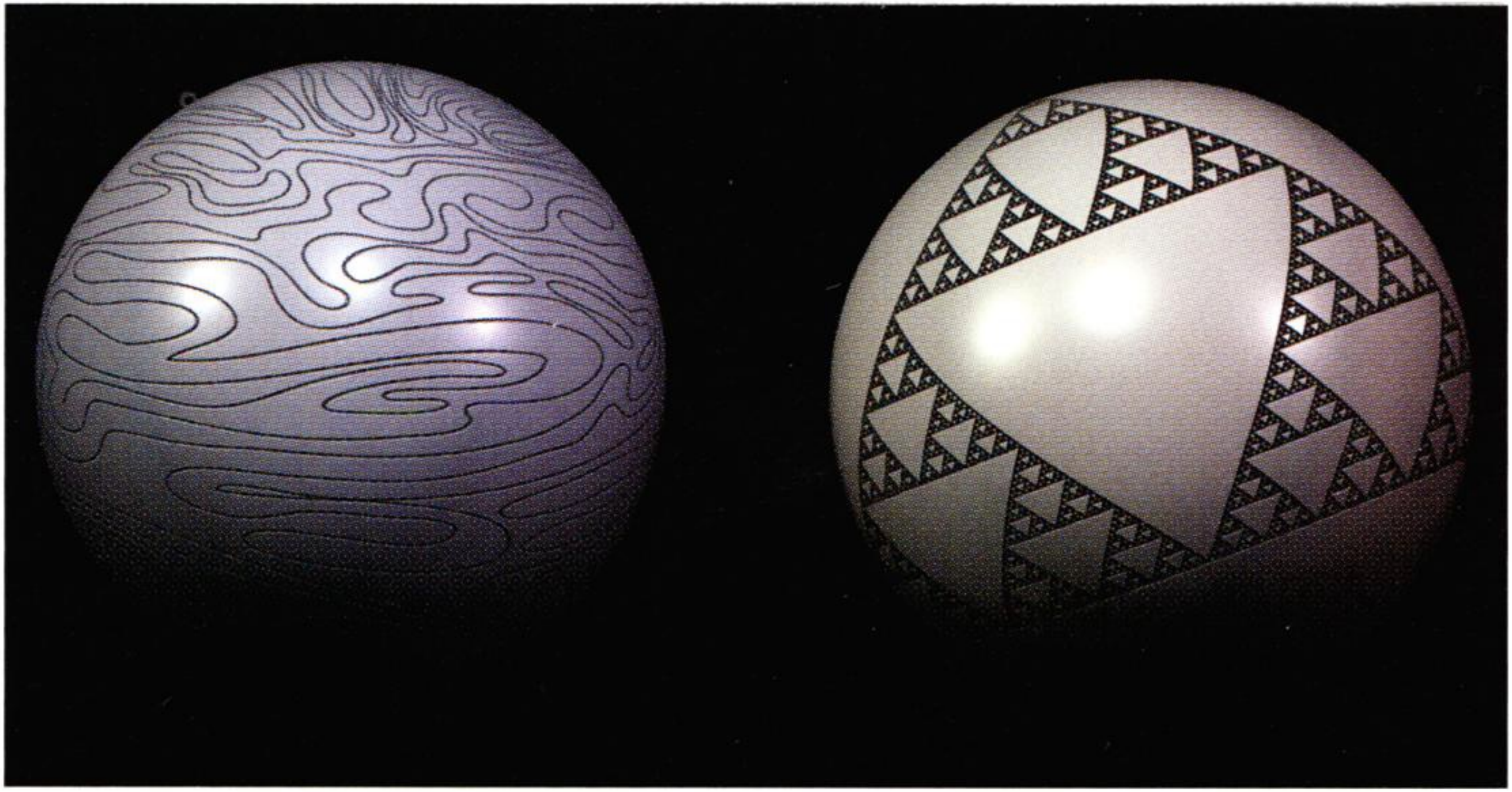
領域を異なるマクロ状態に対応する領域に分解する。



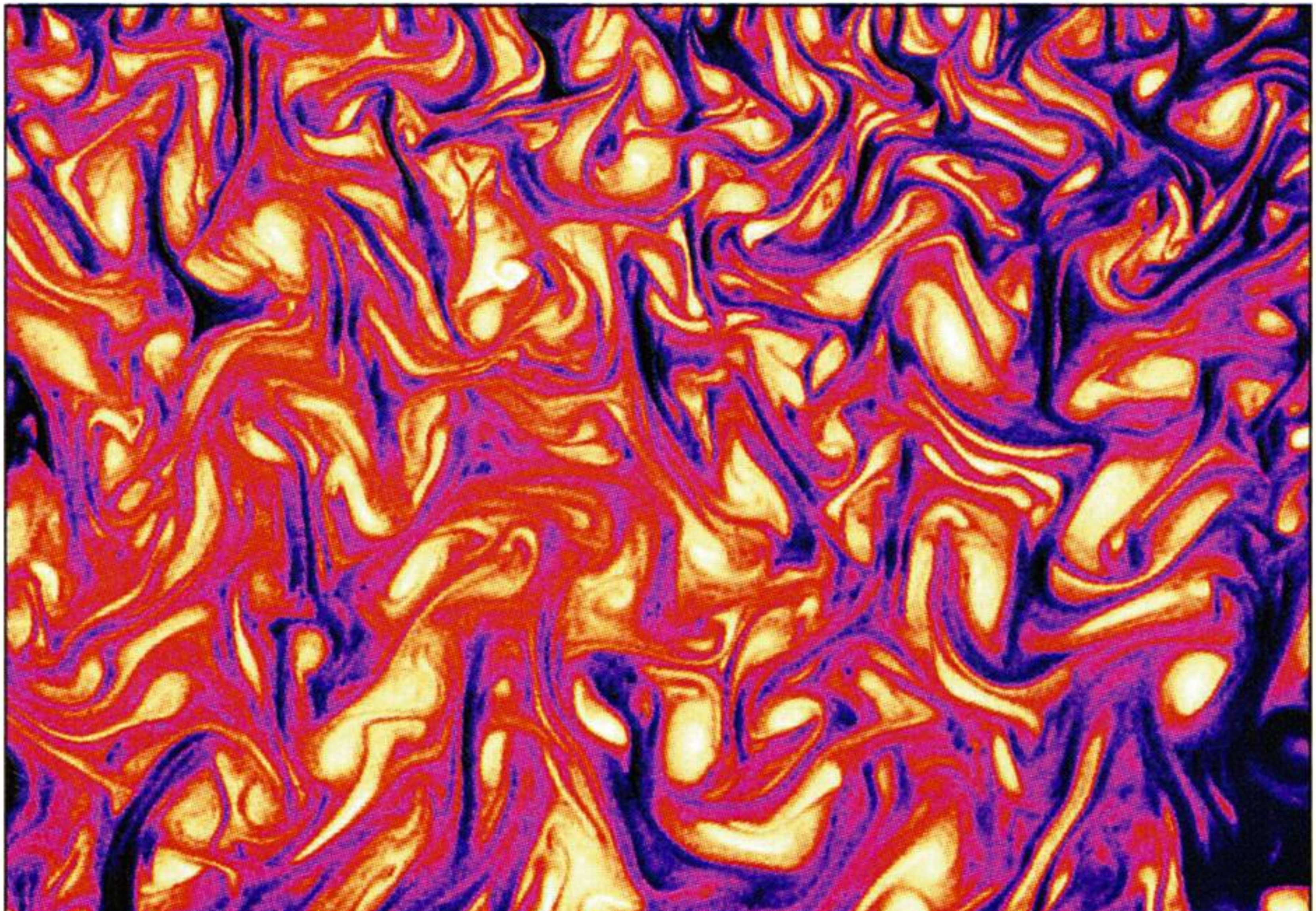


熱平衡と
Coarse-graining

アノマリーが起こり得る境界



fractal



Tsallis エントロピーの性質 2

Tsallis エントロピーは、chain ルールを満たす。

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

Tsallis エントロピーの導出
Baez, Fritz, Leinsterが証明したこと

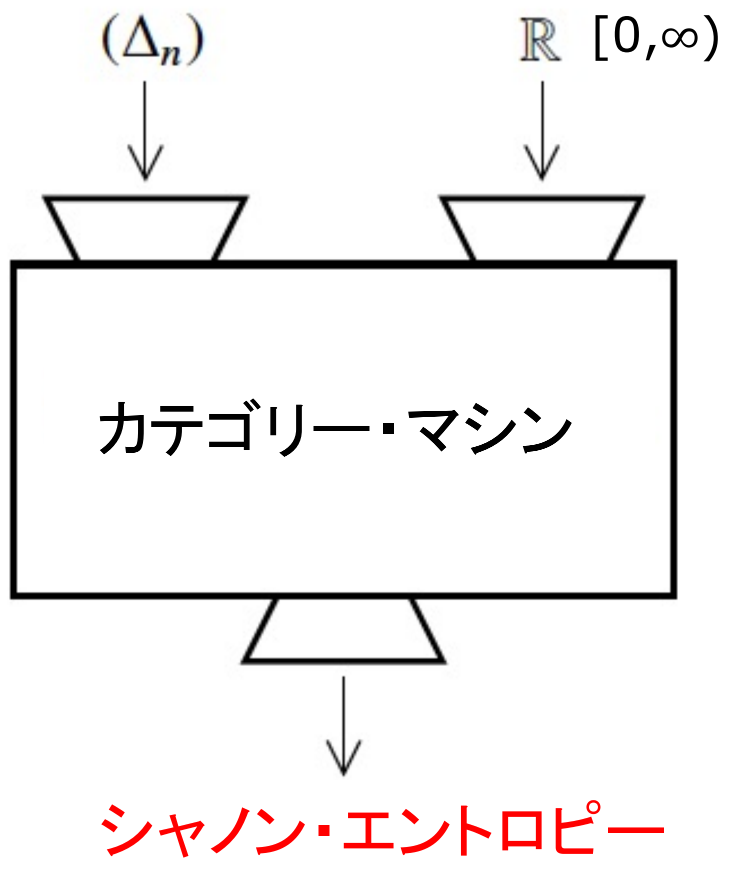
$1 \neq q \in \mathbb{R}$で、$K(f)$は有限な確率空間上の測度を保存する関数fを、実数に写す関数とする。この時、次の二つの条件は同値である。

1. Kは三つの性質を持っている。
   a. すべての同型写像f について $K(f) = 0$
   b. 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。$\mathrm{K}(f \circ g) = \mathrm{K}(f) + \mathrm{K}(g)$
   c. $\mathrm{K}(\lambda f \oplus (1-\lambda) g) = \lambda^q K(f) + (1-\lambda^q) K(g)$
2. ある $c \in \mathbb{R}$が存在して、$K = cL_q$

   ここに、$f: (X, p) \to (Y, s)$の時、$L_q = S_q(p) - S_q(s)$

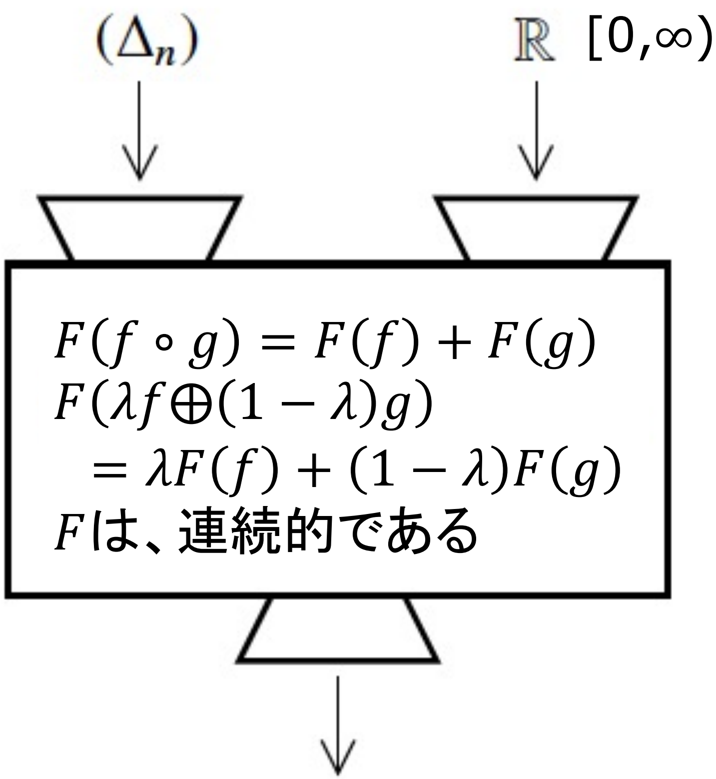
   すなわち、L_q は、Tsallis エントロピーの「情報損失」である。

カテゴリー論の魅力と威力

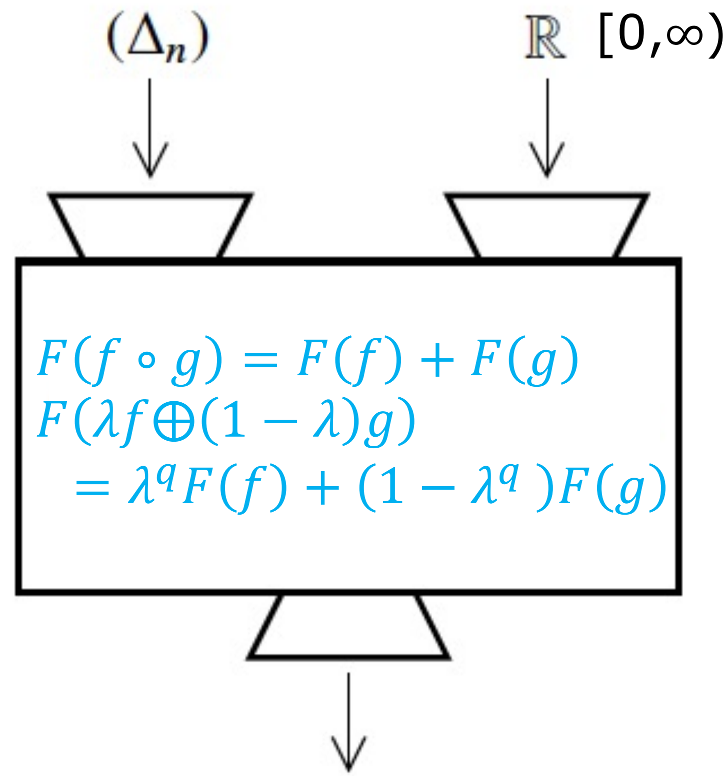


入力に、シンプレックス(Δ_n)と実数軸が与えられれば、
出力にシャノン・エントロピーの概念を返す、一般的な能力を持つカテゴリー論的マシンが存在する！

Tsallis エントロピーの導出



$$H = -\sum_i p_i \log p_i$$



$$S_q = \frac{k}{q-1}\left(1 - \sum_i p_i^q\right)$$

カテゴリーとして実数を捉える

ShannonエントロピーとTsallisエントロピーがFunctorであるという議論を振り返る

ShannonエントロピーとTsallisエントロピーがFunctorであるという議論をしてきたのだが、そこでの重要な論点は、エントロピーが取る非負の実数$[0, \infty)$という値の範囲そのものが、カテゴリーとみなすことができるということである。ただ、実数をカテゴリーとして捉えることができるということは、自明ではない。

確率分布のカテゴリーから $[0, \infty)$というカテゴリーへの、ある性質を満たすFunctor Fは、シャノン・エントロピーを一意に決定するという発見が重要なのだが、ここでは、「実数というカテゴリー」の理解にフォーカスして、その逆の、シャノン・エントロピーの差で表現される「情報の損失」は、Functor とみなせることを、もう少し詳しくみておこう。

一つのオブジェクトしかもたない確率分布 $(\{1\},\{1\})$

一つのオブジェクトしかもたない確率分布$(\{*\},p)$を考えてみよう。これは、確率変数が一つの確率分布$(\{1\},p)$と表してもいい。

一つしか確率変数がないのだから、pは1になる。一つのオブジェクトしかもたない確率分布は、$(\{1\},\{1\})$と表すことができる。

確率分布(X,p)から一つのオブジェクトしかもたない確率分布$(\{1\},\{1\})$への測度を保存する関数を考え、それを $!_p$ と表そう。

確率分布(X, p)から確率分布$(\{1\}, \{1\})$への関数$!_p$

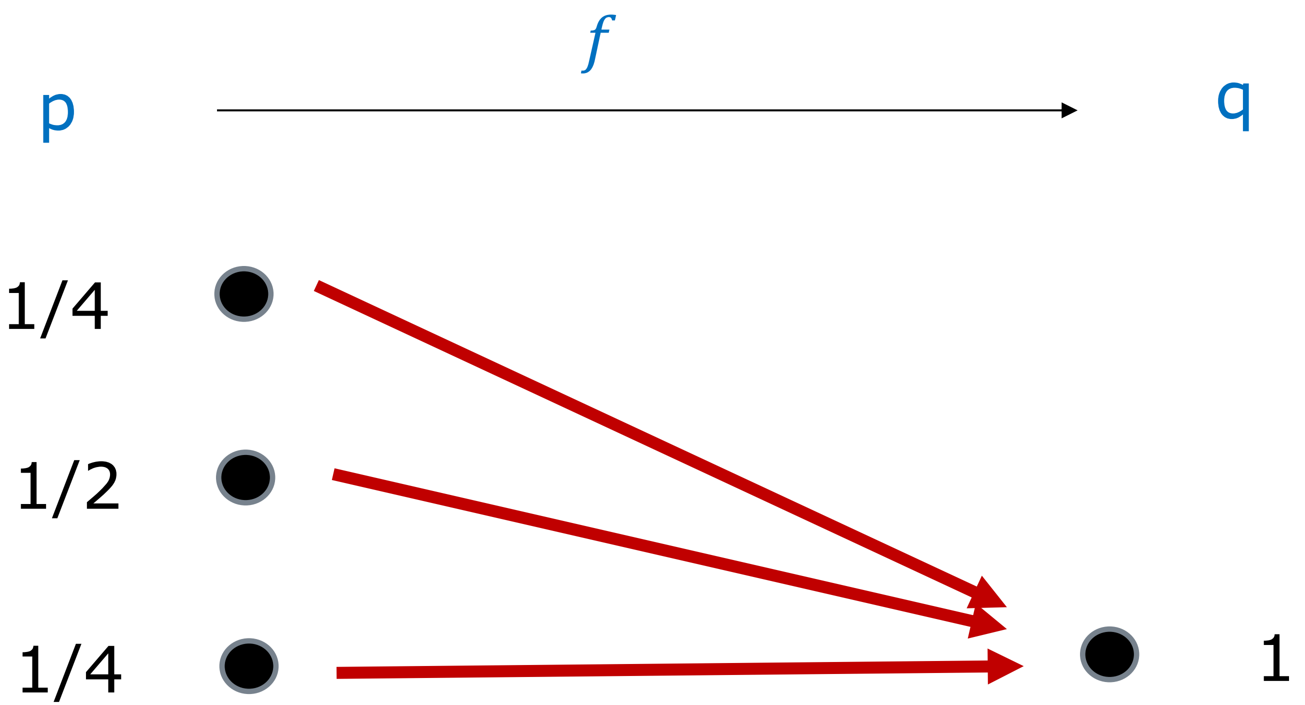
例えば、フェアなコインの確率分布は、({1, 2}, {1/2, 1/2})と表せるのだが、それから ({1}, {1})への関数を考えてみる。

この場合、関数 $!_p$: ({1, 2}, {1/2, 1/2}) → ({1}, {1})は、コインに裏表 {1, 2} があることも、それぞれが等しい確率{1/2, 1/2}を持つことも、すべて忘れることになる。

ある確率分布(X, p)からエントロピーHを得る時、我々は同じことをしている。元の分布が，何個の確率変数を持ち、それぞれがどういう確率を持っていたのかは、捨象され忘れられているから。

関数 $!_p$ の例

次の測度を保存する関数 f は、確率分布(X, p)から$(\{1\}, \{1\})$への関数 $!_p$ の例である。



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

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

測度を保存する関数fと、$!_p$ の関係

測度を保存する任意の関数 f について、次の図式は可換である。

$$
\begin{tikzcd}
(X,p) \arrow[rr, "f"] \arrow[dr, "!_p"'] & & (Y,q) \arrow[dl, "!_q"] \\
& (\{1\},\{1\}) &
\end{tikzcd}
$$

すなわち、$!_q \circ f = !_p$ 。

I_p, I_qも測度を保存する関数である。

cを正の定数、$H(p)$をシャノン・エントロピーとした時、

$$F\left(!_p\right) = cH(p)$$

と定義する。

Fは、測度を保存する関数 $!_p$ を、シャノン・エントロピーの定数倍の実数 $[0, \infty)$に写す関数である。

この時、Fは、次の定義のもとで、カテゴリー$FinProb$から、カテゴリー $[0, \infty)$へのFunctor である。

また、$FinProb$の任意の射 $f\colon p \to q$に対して、次の式が成り立つ。

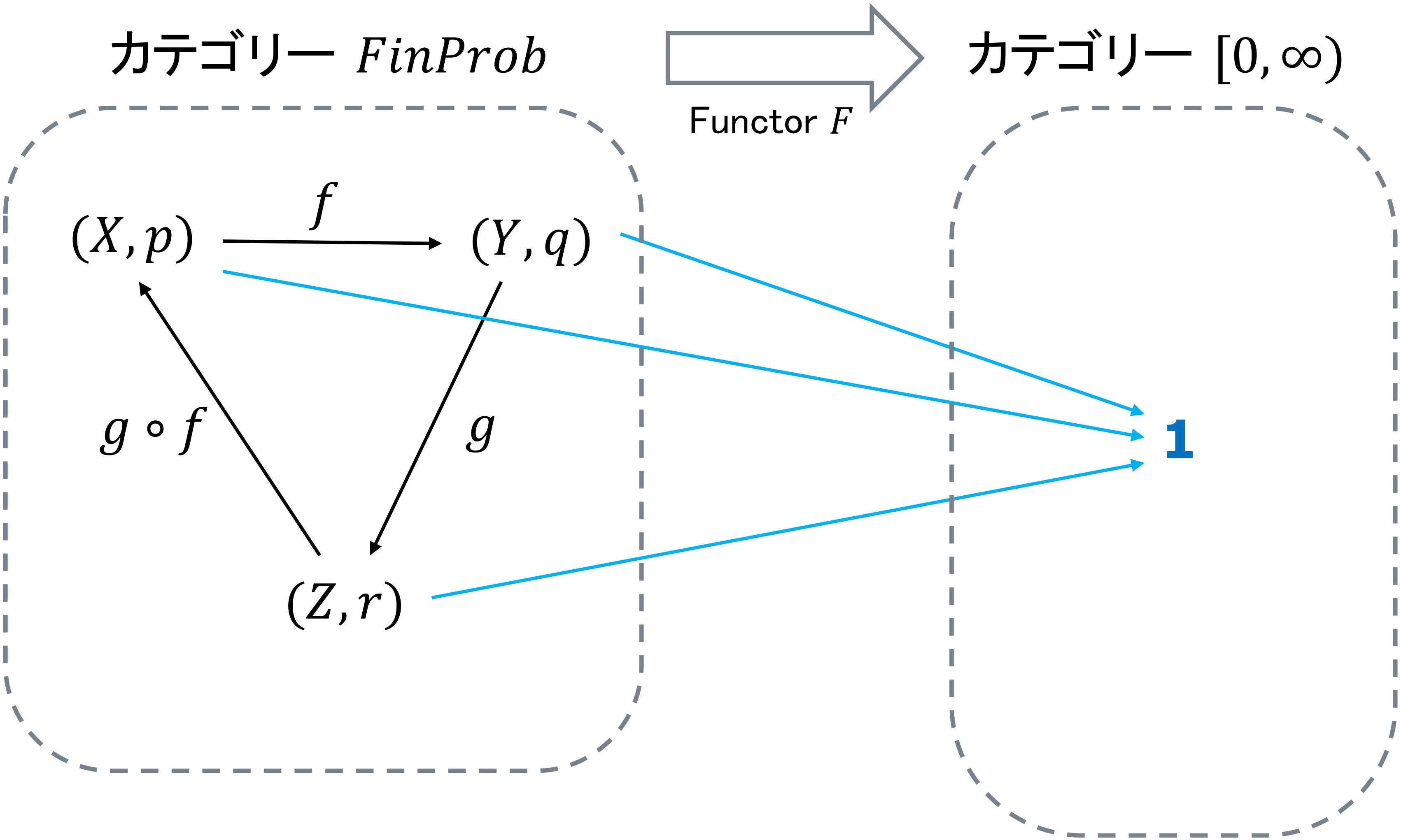
$$F(f) = c(H(p) - H(q))$$

ここで、確率分布のカテゴリー$FinProb$と実数のカテゴリー $[\mathbf{0},\infty)$を次のように定義する

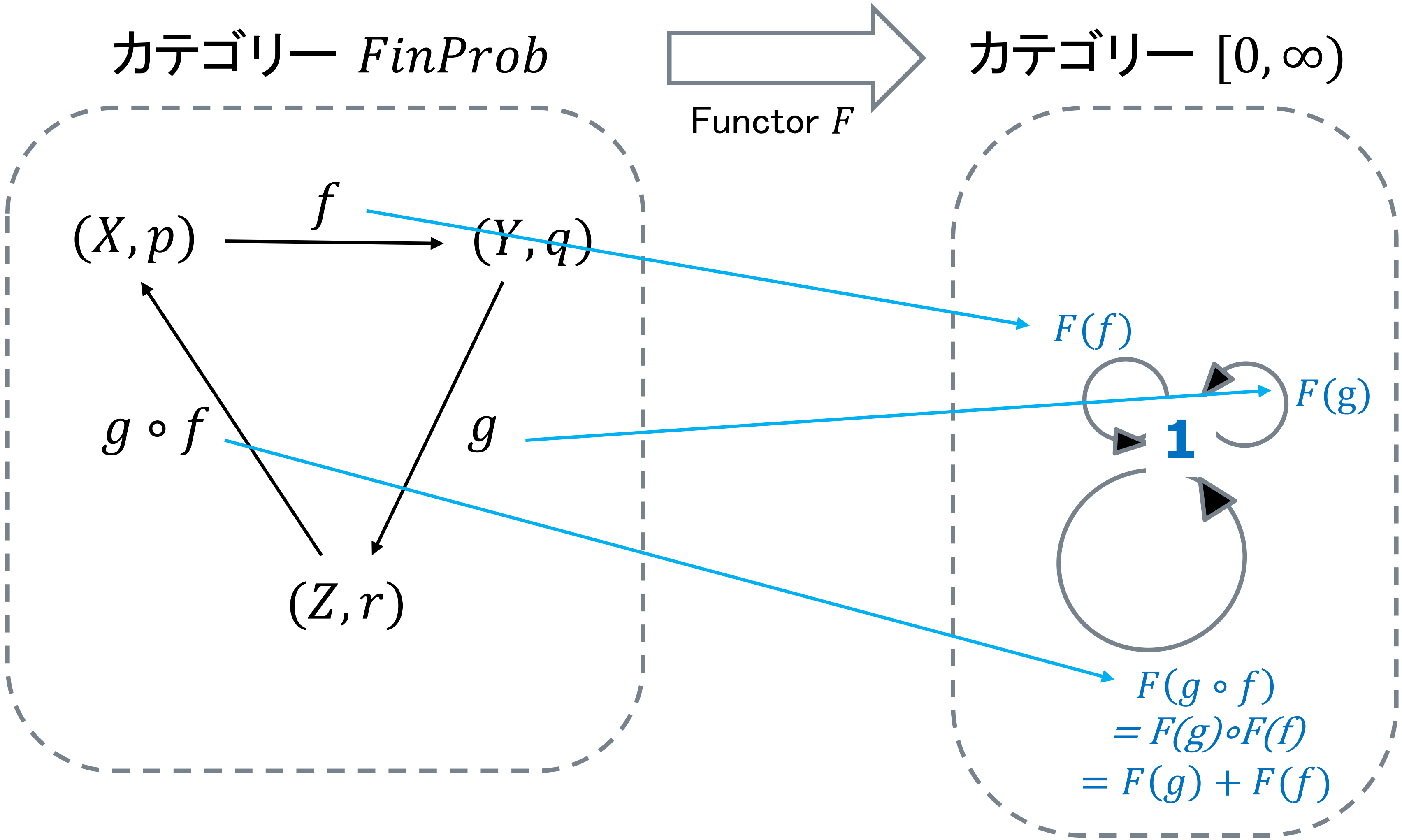
- **FinProb** : 確率分布のカテゴリー
  - オブジェクト : 確率分布 $(\mathrm{X},\mathrm{p}),(\mathrm{Y},\mathrm{q}),(\mathrm{Z},\mathrm{r}),\ldots$
  - 射 : 測度を保存する関数 $f:(X,p)\to(Y,q), g:(Y,q)\to(Z,r),\ldots$
  - 射の合成 : 関数の合成 $f\circ g,\ldots$

- $[\mathbf{0},\infty)$: 実数のカテゴリー
  - オブジェクト : 実数カテゴリーは一つのオブジェクトしか持たない
    これを$*$とか1とかで表す。
  - 射 : $*\to*$
    実数 $r,s,\ldots$ は、射 $*\to*$ で表現される
  - 射の合成 :
    射$r:*\to*$と射$s:*\to*$ の合成は、$r+s:*\to*$として和で定義される

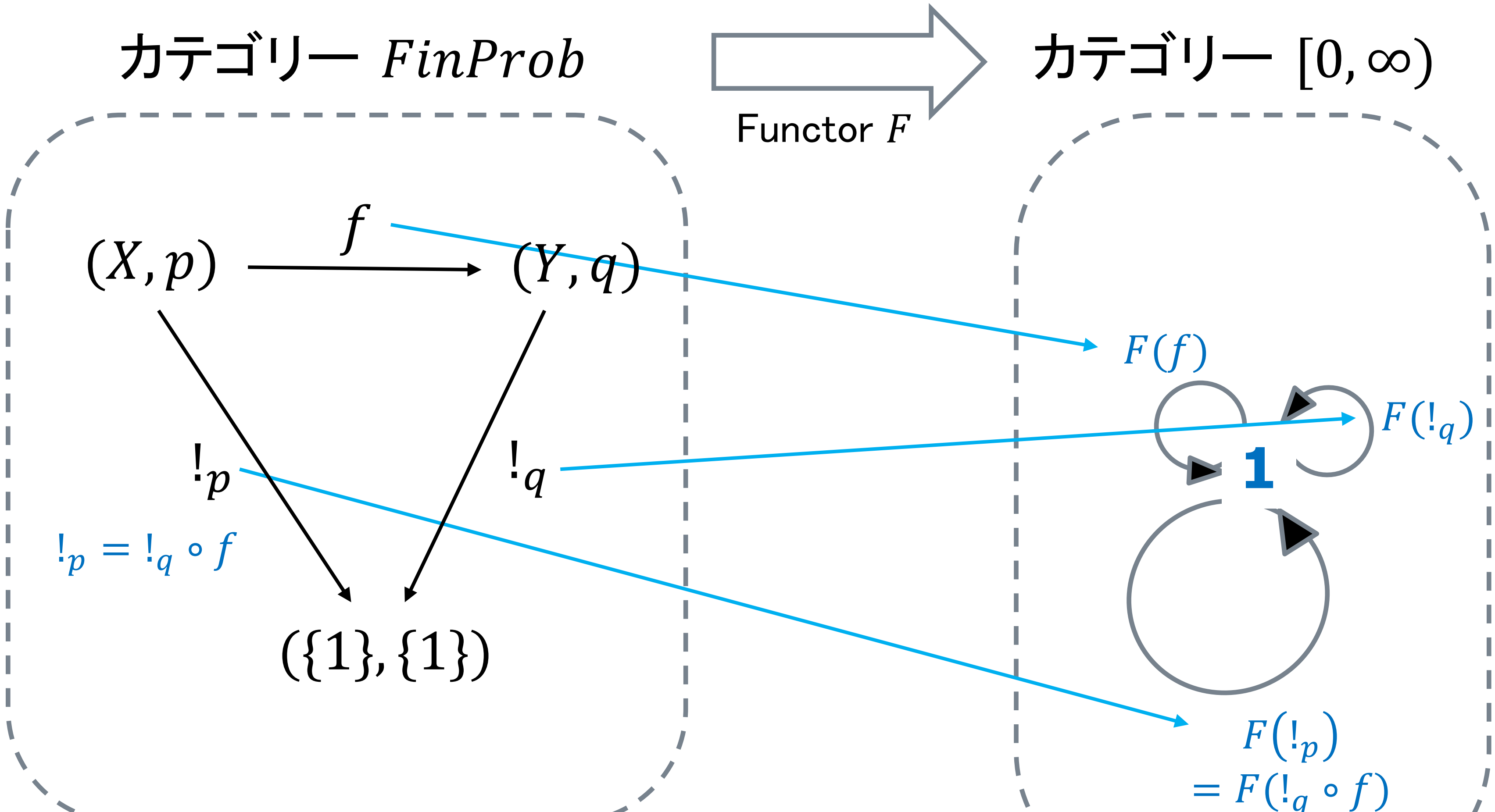
$FinProb$から$[0,\infty)$へのFunctor F オブジェクトへのFの作用



$FinProb$から$[\mathbf{0},\infty)$へのFunctor F
射への作用



$FinProb$から$[\mathbf{0},\infty)$へのFunctor F
射 $f, !_p, !_q$への作用



$F(f) = cH(p) - cH(q) \Leftarrow F(f) = F(!_p) - F(!_q) \Leftarrow$

実数 $[0,\infty)$をカテゴリーと見るアイデアの起源

実数 $[0,\infty)$をカテゴリーと見るアイデアは、1973年のLawvereの論文 “Metric spaces, generalized logic and closed categories” によるものである。
http://www.tac.mta.ca/tac/reprints/articles/1/tr1.pdf

X	hom-values for X	composition law and identity law for X	domain of composition law for X	domain of identity law for X
metric space	nonnegative real quantities	$\geqq$	sum	zero
category	abstract sets	mapping	cartesian product	one element set

q-*logarithmic*エントロピー

Tsallis エントロピー = q-対数エントロピー

Tsallis エントロピー を、「q-対数エントロピー (q-logarithmic entropies)」ということがある。

Tsallis エントロピー は、Shannonエントロピーの定義中の log をq-対数 $\ln_q$ に置き換えたものである。

Shannon エントロピー $\boldsymbol{H}$

$$H = \sum_i p_i \log\left(\frac{1}{p_i}\right)$$

$\boldsymbol{Tsallis\ (q{-}logarithmic)}$ エントロピー $\boldsymbol{S_q}$

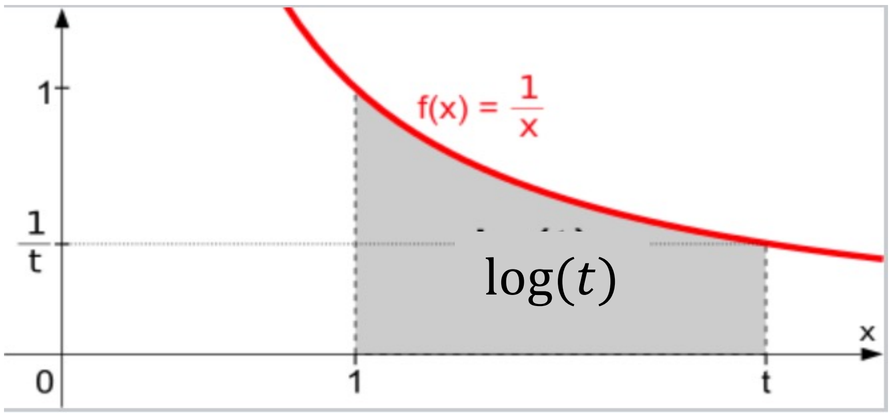
$$S_q = \sum_i p_i \ln_q\left(\frac{1}{p_i}\right)$$

t^nの積分とlog

$n \neq -1$の時　　$$\int t^n \, dt = \frac{t^{n+1}}{n+1} + C$$

$n = -1$の時　　$$\int t^{-1} \, dt = \log(t) + C$$

$$\ln(x) = \log_e(x) = \int_1^x t^{-1} \, dt$$



q-logarithm $\ln_q$ を計算する

$q \neq 1$の時 $$\int t^{-q}\,dt = \frac{t^{-q+1}}{-q+1} + C$$

$$\ln_q(x) = \int_1^x t^{-q}\,dt$$ とする。

$$\ln_q(x) = \left[\frac{t^{-q+1}}{-q+1}\right]_1^x = \frac{x^{-q+1}}{-q+1} - \frac{1^{-q+1}}{-q+1}$$

$$= \frac{x^{1-q}-1}{1-q} = \frac{1-x^{1-q}}{q-1}$$

$\ln$ と $\ln_q$

$$\ln(x) = \int_1^x t^{-1}\,dt$$

$$\ln_{\mathrm{q}}(x) = \int_1^x t^{-q}\,dt$$

$q \neq 1$ の時　$$\ln_{\mathrm{q}}(x) = \frac{x^{1-q} - 1}{1-q}$$

$$\ln_{\mathrm{q}}\left(\frac{1}{x}\right) = \frac{x^{q-1} - 1}{1-q}$$

$q = 1$ の時　$\ln_1(x) = \log(x)$

Shannon エントロピー H と $Tsallis$ (q−$logarithmic$) エントロピー S_q

Shannon エントロピー H

$$H = \sum_i p_i \, log\left(\frac{1}{p_i}\right)$$

$Tsallis$ (q−$logarithmic$) エントロピー S_q

$$S_q = \sum_i p_i \ln_{\mathrm{q}}\left(\frac{1}{p_i}\right)$$

$Tsallis\ (q\text{-}logarithmic)$ エントロピー S_q

$$S_q = \sum_i p_i \ln_q \left(\frac{1}{p_i}\right)$$

$$= \sum_i p_i \left(\frac{1 - p_i^{q-1}}{q-1}\right)$$

$$= \frac{1}{q-1}\left(\sum_i p_i - \sum_i p_i^{q-1+1}\right)$$

$$= \frac{1}{q-1}\left(1 - \sum_i p_i^q\right)$$

$$S_q = \frac{1}{q-1}\left(1 - \sum_i p_i^q\right)$$

‘q-logarithmic entropy’ と呼ぶべきだという主張

The q-logarithmic entropies have been discovered and rediscovered repeatedly. They seem to have first appeared in a 1967 paper on information and classification by Havrda and Charv´at [139] …

Further work on the entropies was carried out in 1968 by Vajda [337] (with reference to Havrda and Charv´at). They were rediscovered in 1970 by Dar´oczy [76] (without reference to Havrda and Charv´at), and were the subject of Section 6.3 of the 1975 book [3] by Acz´el and Dar´oczy (with reference to all of the above). …

‘q-logarithmic entropy’ と呼ぶべきだという主張

In physics, meanwhile, the q-logarithmic entropies appeared in a 1971 article of Lindhard and Nielsen [227] (according to Csiszar [73], Section 2.4).

They also made a brief appearance in a review article on entropy in physics by Wehrl ([347], p. 247). Finally, they were rediscovered again in a 1988 paper on statistical physics by Tsallis [328] (with reference to none of the above).

T. Leinster, Entropy and Diversity,
p. 95 Remark 4.1.4
https://arxiv.org/pdf/2012.02113.pdf

RényiエントロピーとHill数

Rényi エントロピー　$\mathrm{H_q}$

Rényi エントロピー H_qは、次の式で定義される。

$$H_q(\boldsymbol{p}) = \frac{1}{1-q} \log \left(\sum_i p_i^q \right)$$

先に見た Tsallis エントロピー S_q とは、関係がある。

$$S_q(\boldsymbol{p}) = \frac{1}{q-1} \left(1 - \sum_i p_i^q \right)$$

どちらも

$X = \sum_i p_i^q$ の関数とみなせる。

Rényi エントロピー $\mathrm{H_q}$ と Tsallis エントロピー S_q の関係

$X = \sum_i p_i^q$ とする。

$$H_q(\boldsymbol{p}) = \frac{1}{1-q}\log\left(\sum_i p_i^q\right) = \frac{1}{1-q}\log(X)$$

$$(1-q)H_q(\boldsymbol{p}) = \log(X)$$

$$\exp\Big((1-q)H_q(\boldsymbol{p})\Big) = X$$

$$S_q(\boldsymbol{p}) = \frac{1}{q-1}\left(1-\sum_i p_i^q\right) = \frac{1}{q-1}(1-X)$$

$$(q-1)S_q(\boldsymbol{p}) = 1-X$$

$$X = 1+(1-q)S_q(\boldsymbol{p})$$

$$\therefore \exp\Big((1-q)H_q(\boldsymbol{p})\Big) = 1+(1-q)S_q(p)$$

Rényi エントロピー $\mathrm{H_q}$ と Tsallis エントロピー S_q の関係

$$\exp\left((1-q)H_q(\boldsymbol{p})\right) = 1 + (1-q)S_q(\boldsymbol{p})$$

$$(1-q)H_q(\boldsymbol{p}) = \log(1 + (1-q)S_q(\boldsymbol{p}))$$

$$H_q(\boldsymbol{p}) = \frac{1}{1-q}\log((1-q)S_q(\boldsymbol{p}) + 1)$$

$$\exp\left((1-q)H_q(\boldsymbol{p})\right) = 1 + (1-q)S_q(\boldsymbol{p})$$

$$\exp((1-q)H_q(\boldsymbol{p})) - 1 = (1-q)S_q(\boldsymbol{p})$$

$$S_q(\boldsymbol{p}) = \frac{1}{1-q}(\exp((1-q)H_q(\boldsymbol{p})) - 1)$$

$q \neq 1$ の時　$\ln_{\mathrm{q}}(x) = \dfrac{x^{1-q} - 1}{1 - q}$

$$
\begin{aligned}
\ln_q(\exp H_q(\boldsymbol{p})) &= \frac{1}{1-q}\left(\exp\left(H_q(\boldsymbol{p})\right)^{1-q} - 1\right) \\
&= \frac{1}{1-q}\left(\left(\exp(1-q)\, H_q(p)\right) - 1\right) \\
&= S_q(\boldsymbol{p})
\end{aligned}
$$

$$\ln_q(\exp H_q(\boldsymbol{p})) = S_q(\boldsymbol{p})$$

$q \neq 1$ の時　$\exp_{\mathrm{q}}(x) = (1 + (1-q)x)^{1/(1-q)}$とする。

$$\ln_q \exp_q(x) = \frac{\exp_q(x)^{1-q} - 1}{1-q} = \frac{(1+(1-q)x) - 1}{1-q} = x$$

$$\log\left(\exp_q S_q(\boldsymbol{p})\right) = \log\left(1 + (1-q)S_q(\boldsymbol{p})\right)^{1/(1-q)}$$
$$= \frac{1}{1-q}\log\Big((1-q)S_q(\boldsymbol{p}) + 1\Big) = H_q(\boldsymbol{p})$$

$$log(exp_q S_q(\boldsymbol{p})) = \mathrm{H}_{\mathrm{q}}(\mathbf{p})$$

$q \neq 1$ の時　$\ln_{\mathrm{q}}(x) = \dfrac{x^{1-q} - 1}{1 - q}$

$q \neq 1$ の時　$\exp_{\mathrm{q}}(x) = (1 + (1 - q)x)^{1/(1-q)}$

とすると、

$$\ln_q \exp_q(x) = x$$

この時

$$S_q(\boldsymbol{p}) = \ln_q(\exp H_q(\boldsymbol{p}))$$

$$\mathrm{H_q}(\mathbf{p}) = log(exp_q S_q(\boldsymbol{p}))$$

Hill number D_q

$$D_q(\boldsymbol{p}) = \exp H_q(\boldsymbol{p})$$

を、次数 q のHill数 という。

Hill数 $D_q(\boldsymbol{p})$を使うと、Rényi エントロピー $\mathrm{H_q}(\boldsymbol{p})$、Tsallis エントロピー $S_q(\boldsymbol{p})$は、次のように表すことができる。

$$H_q(\boldsymbol{p}) = \log D_q(\boldsymbol{p})$$

$$S_q(\boldsymbol{p}) = \ln_q D_q(\boldsymbol{p})$$

拡大されたエントロピー概念

$$\boldsymbol{Shannon, Boltzmann, Gibbs}$$

$$H = -\sum_i p_i \log p_i$$

$$\boldsymbol{Tsallis\ (q - logarithmic)}$$

$$S_q = \frac{1}{q-1}\left(1 - \sum_i p_i^q\right)$$

Rényi

$$H_q = \frac{1}{1-q}\log\left(\sum_i p_i^q\right)$$

拡大されたエントロピー概念

$\boldsymbol{Shannon, Boltzmann, Gibbs}$

$$H = -\sum_i p_i \log p_i$$

$\boldsymbol{Hill\ number}$

$$D_q = \exp H_q$$

$\boldsymbol{Tsallis\ (q - logarithmic)}$

$$S_q = \ln_q D_q$$

$\boldsymbol{Rényi}$

$$H_q = \log D_q$$