

CATEGORIES OF CONVEX SETS AND OF METRIC SPACES,  
WITH APPLICATIONS TO STOCHASTIC PROGRAMMING  
AND RELATED AREAS

Xiao-qing Meng

SEL  
DISS  
M3  
1988  
M46

A dissertation submitted to the Faculty of the Graduate School  
of State University of New York at Buffalo in partial fulfillment  
of the requirements for the degree of Doctor of Philosophy

NYB489-B17456

6/16

## ABSTRACT

Let  $X$  and  $Y$  be two measurable spaces with  $\sigma$ -algebra  $\mathcal{A}$  and  $\mathcal{B}$  respectively. A random map  $f: X \rightarrow Y$  is a function  $f: X \times \mathcal{B} \rightarrow [0, 1]$ , such that 1).  $f(x, \_)$  is a probability measure on  $Y$ , for each  $x \in X$ ; and 2).  $f(\_, B)$  is a  $\mathcal{A}$ -measurable function for each  $B \in \mathcal{B}$ . The measurable spaces form the objects of two categories,  $\mathcal{M}$  and  $\mathcal{P}(\mathcal{M})$ . The first the morphisms are simply the measurable functions; in the second they are the random maps.  $\mathcal{P}(\mathcal{M})$  provides a base ground on which one can formulate, analyse and solve many problems in probability theory, statistical decision theory, stochastic processes and stochastic dynamic programming. Further more,  $\mathcal{P}(\mathcal{M})(X, Y)$ , the space of random maps between  $X$  and  $Y$  has various structures: it is a convex set, it is also a generalized metric space. This actually induced by a closed functor  $\varphi$  from  $\mathcal{C}$ , the category of convex sets and affine linear maps, to  $\mathcal{V}\text{-Cat}$ , the category of generalized metric spaces and distance decreasing functions. Then we will use this generalized metric on  $\mathcal{P}(\mathcal{M})(X, Y)$  to give criterions of best decision rules and best optimal policies in the statistical decision problems and stochastic dynamic programming problems.

## CONTENT

Introduction.

Chapter 1. The category of convex sets  $\mathbf{C}$ .

1.  $\mathbf{C}$  is a closed category 1.1
2.  $\mathbf{C}$  is a coreflective subcategory of  $K\text{-mod}/K$  1.8
3. A closed functor  $\varphi$  from  $\mathbf{C}$  to  $\mathcal{V}\text{-Cat}$  1.21
4.  $\varphi$  has a left adjoint 1.26
5. The relation between posets and  $\mathcal{V}$ -categories 1.29

Chapter 2. The category  $\mathcal{P}(\mathcal{A})$ .

1. The category  $\mathcal{P}(\mathcal{A})$  and its properties 2.1
2. The covariant and contravariant topologies  
of  $\mathcal{P}(\mathcal{A})(X, Y)$  2.14

Chapter 3. Applications in statistics, stochastic processes  
and stochastic dynamic programming.

1. Statistical decision problems 3.1
2. Stochastic processes, transformation and dynamics 3.12
3. Decomposition of stochastic transformations 3.33
4. Stochastic dynamic programming 3.45

## INTRODUCTION

Let  $X$  and  $Y$  be two measurable spaces with  $\sigma$ -algebras  $\mathcal{A}$  and  $\mathcal{B}$  respectively. A random map (i.e. probabilistic transformation)  $f$  from  $X$  to  $Y$  is a map  $f: X \times \mathcal{B} \rightarrow [0,1]$ , such that for each  $x \in X$ ,  $f(x, \cdot)$  is a probability measure on  $Y$  and for each  $B \in \mathcal{B}$ ,  $f(\cdot, B)$  is  $\mathcal{A}$ -measurable. The measurable spaces form the objects of two categories,  $\mathcal{M}$  and  $\mathcal{P}(\mathcal{M})$ . The first the morphisms are simply the measurable functions; in the second they are the random maps.  $\mathcal{P}(\mathcal{M})$  provides a base ground on which one can formulate, analyse and solve many problems in probability theory, statistical decision theory and stochastic processes. This fact has been already pointed out by Cencov [4], Lawvere [10] and Schiopu [13].

Since for each pair of measurable spaces  $X$  and  $Y$ , the set  $\mathcal{P}(\mathcal{M})(X, Y)$  of random maps between  $X$  and  $Y$  is a convex set, we begin this work with the study of the category  $\mathcal{C}$  of convex sets and affine linear maps. In chapter 1, we show that  $\mathcal{C}$  is a closed category and there is a closed functor  $\varphi$  from  $\mathcal{C}$  to the category  $\mathcal{U}$ -cat of generalized metric spaces. Therefore  $\mathcal{P}(\mathcal{M})$  is enriched in  $\mathcal{C}$ , and, so, is enriched in  $\mathcal{U}$ -cat. This gives to each convex set, and specifically to  $\mathcal{P}(\mathcal{M})(X, Y)$ , a generalized metric structure. We write  $K$  for the semi-ring of nonnegative real numbers and prove that  $\mathcal{C}$  is a coreflective subcategory of  $K\text{-mod}/K$ . Moreover,  $\varphi$  has a left adjoint which assigns to each generalized metric space the free convex set it generates.

In chapter 2, we study the category  $\mathcal{P}(\mathcal{M})$  and the topologies induced by the above generalized metric on  $\mathcal{P}(\mathcal{M})(X, Y)$ .

For each measurable space  $X$  with  $\sigma$ -algebra  $\mathcal{A}$ , let  $\mathcal{P}(X)$  be the set of probability measures on  $X$  with the smallest  $\sigma$ -algebra such that for each  $A \in \mathcal{A}$ , the evaluation map  $ev_A: \mathcal{P}(X) \rightarrow [0,1]$ ,  $ev_A(p) = p(A)$  is measurable; for each measurable function  $f: X \rightarrow Y$ , there is a measurable function  $\mathcal{P}(f): \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$ , defined by  $\mathcal{P}(f)(p)(B) = p(f^{-1}(B))$ , for  $p \in \mathcal{P}(X)$ ,  $B \in \mathcal{B}$ . In this way,  $\mathcal{P}$  forms a monad on  $\mathcal{M}$ , and  $\mathcal{P}(m)$  is just the Kleisli category of it. On each convex set  $C$  there are two different topologies induced by the generalized metric: a contravariant one with basis  $\{B(a,r) = \{c \in C \mid \text{dist}(a,c) < r\} \mid a \in C, r \in K\}$ , and a covariant one with basis  $\{B(a,r) = \{c \in C \mid \text{dist}(a,c) < r\} \mid a \in C, r \in K\}$ , where we use  $\text{dist}$  to denote the generalized metric. Since this metric  $\text{dist}$  is not symmetric, these two topologies are quite different than the topology induced by the usual symmetric metric. However when  $C = \mathcal{P}(m)(X,Y)$ , they are well-behaved: both topologies are Hausdorff and sequentially Cauchy complete under three different definitions of Cauchy sequences. Also, the contravariant topology is compact if  $X$  and  $Y$  are finite.

In chapter 3, we study certain applications of  $\mathcal{P}(m)$  and this generalized metric on  $\mathcal{P}(m)(X,Y)$  the set of random maps between each pair of measurable spaces  $X$  and  $Y$ . Using random maps to formulate the concepts of stochastic processes, dynamic stochastic programming and statistical decision problems clarifies these concepts.

The usual formulations of stochastic processes included two

different types. One emphasizes the results of the dynamics by defining a stochastic process as a sequence of probability measures on each accumulated state space  $\prod_{k \leq n} X_k$ , compatible with the projections  $\pi_{(n)} : \prod_{k \leq n+1} X_k \rightarrow \prod_{k \leq n} X_k$ , that is the following triangle is commutative.

$$\begin{array}{ccc}
 1 & \xrightarrow{p_{n+1}} & \prod_{k \leq n+1} X_k \\
 & \searrow p_n & \downarrow \pi_{(n)} \\
 & & \prod_{k \leq n} X_k
 \end{array}$$

The other one emphasizes the dynamics itself, namely it gives the dynamics  $q_n : \prod_{k \leq n} X_k \rightarrow X_{n+1}$  at each stage  $n$ . To distinguish the latter one we call it stochastic dynamics.

Given a stochastic dynamics, it is easy to get its results, namely the corresponding stochastic process, but the other way around may not be possible. A Markov dynamics is a stochastic dynamics such that the dynamics at each stage is independent of history, i.e.  $q_n : X_n \rightarrow X_{n+1}$ . There is an adjoint relation between the category of stochastic dynamics and the category of Markov dynamics.

The generalized metric provides a possible criterion to choose the "best" decision rules (in statistical decision problems) and policies (in multistage decision problems or dynamic stochastic programming). We pursue this in chapter 3. For example, in case of the statistical decision problem,  $X$  is the message space,  $\Omega$  is the observation space and  $Y$  is the decision space. A random map

$f: X \rightarrow \Omega$  describes the process of observation through a noisy channel;  $g: X \rightarrow Y$  is the correct action which reflects the objective connection between message and decisions ( $g$  also could be random). For any decision rule  $\delta: \Omega \rightarrow Y$ , we write  $\text{dist}(\delta \circ f, g)$  for the distance between  $\delta \circ f$  and  $g$ , then  $\text{dist}(\delta \circ f, g)$  is a measure of  $\delta$  (with respect to  $f$  and  $g$ ). An optimal decision rule  $\delta^*$  should satisfy the property

$$\text{dist}(\delta^* \circ f, g) = \inf_{\delta} \text{dist}(\delta \circ f, g)$$

Note that,  $\text{dist}(\cdot \circ f, g): \mathcal{P}(m)(\Omega, Y) \rightarrow [0, \infty]$  is a contracting (i.e. Lipschitz) function with constant 1, so it is continuous. We show the existence of such optimal decision rules or policies.

CHAPTER 1. THE CATEGORY OF CONVEX SETS  $\mathcal{C}$ 

In this chapter, we are going to discuss the the properties of the category of convex sets  $\mathcal{C}$ , which are the following: (1).  $\mathcal{C}$  is a closed category, (2).  $\mathcal{C}$  is equivalent to a coreflective subcategory of  $K\text{-mod}/K$ , (3). there is a closed functor  $\varphi$  from  $\mathcal{C}$  to  $\mathcal{U}\text{-cat}$ , the category of generalized metric spaces, (4). functor  $\varphi$  has a left adjoint, and some other properties related with  $\mathcal{C}$ .

§1. The category of convex sets  $\mathcal{C}$  is a closed category.

Let  $K = \mathbb{R}^+ = \{r \in \mathbb{R}, \text{ the reals } \mid r \geq 0\}$ .

Definition 1. The theory of  $K$ -module is a category  $\mathbb{T}$  with finite products and a distinguished object  $A$ , such that every object in  $\mathbb{T}$  is one of the sequence

$$A^0 = 1,$$

$$A^{n+1} = A^n \times A, n = 0, 1, 2, \dots$$

Morphisms are

$$0\text{-ary operation } 0: 1 \longrightarrow A,$$

$$\text{unary operation } r: A \longrightarrow A, \text{ for each } r \in K,$$

$$\text{binary operation (addition) } +: A \times A \longrightarrow A,$$

such that

(0).  $1_A = 1_K$ , where  $1_A$  is the identity of  $A$ ,  $1_K$  is the unit of the multiplication in  $K$ ;

and the following diagrams are commutative:

(1). commutativity of  $+$ :

$$\begin{array}{ccc}
 A \times A & \xrightarrow{+} & A \\
 \downarrow (\pi_1, \pi_2) & \searrow & \uparrow \\
 A \times A & & A
 \end{array}$$

where  $\pi_i: A \times A \rightarrow A$ ,  $i = 1, 2$  are the projections;

(2). associativity of +:

$$\begin{array}{ccc}
 (A \times A) \times A & \approx & A \times (A \times A) \\
 \downarrow + \times 1_A & & \downarrow 1_A \times + \\
 A \times A & & A \times A \\
 \searrow + & & \swarrow + \\
 & A &
 \end{array}$$

(3). "0 + a = a":

$$\begin{array}{ccc}
 1 \times A & \xrightarrow{0 \times 1_A} & A \times A \\
 \searrow \cong & & \downarrow + \\
 & & A
 \end{array}$$

(4). "0 · a = 0":

$$\begin{array}{ccc}
 A & \xrightarrow{0} & A \\
 \downarrow 1 & \searrow 0 & \uparrow \\
 & & A
 \end{array}$$

for  $A \xrightarrow{0} A$ , 0 is in K;

(5). " $r \cdot a + r' \cdot a = (r + r') \cdot a$ ":

$$\begin{array}{ccccc}
 A & \xrightarrow{\Delta} & A \times A & \xrightarrow{r \times r'} & A \times A \\
 & \searrow & & & \downarrow + \\
 & & & & A \\
 & & & \swarrow r+r' &
 \end{array}$$

where  $\Delta$  is the diagonal map;

(6). " $r \cdot (r' \cdot a) = (rr') \cdot a$ ":

$$\begin{array}{ccc}
 A & \xrightarrow{r'} & A \\
 \searrow rr' & & \downarrow r \\
 & & A
 \end{array}$$

Definition 2. The theory of convex set is just the affine part  $\mathbb{A}$  of the theory  $\mathbb{T}$  (see J. Isbell, M. Klun and S. Schanuel, Affine part of algebraic theories). So,  $\mathbb{A}$  is a subcategory of  $\mathbb{T}$  with the same objects but only the idempotent operations. An operation  $u: A^n \rightarrow A$  is idempotent, if  $u \circ \Delta_n = 1_A$ , where  $\Delta_n$  is the diagonal from  $A$  to  $A^n$ . Since the  $n$ -ary operations in  $\mathbb{T}$  have the form  $\Sigma_n^\circ(r_1, \dots, r_n)$ , where  $r_i \in K$ ,  $i=1, \dots, n$ ;  $\Sigma$  is defined inductively,  $\Sigma_1 = 1_A$ ,  $\Sigma_2 = +$ ,  $\Sigma_{n+1} = + \circ \Sigma_n \times 1_A: A^{n+1} \approx A^n \times A \xrightarrow{\Sigma_n \times 1_A} A \times A \xrightarrow{+} A$ , therefore,  $\Sigma_n^\circ(r_1, \dots, r_n)$  is idempotent if and only if  $\sum_{i=1}^n r_i = 1$ .

Definition 3. A  $K$ -module  $M$  (a model of theory  $\mathbb{T}$  in  $S$ ,  $S$  is the category of sets) is a finite product preserving functor from  $\mathbb{T}$  to  $S$ .

Let  $S^{\mathbb{T}}$  be the category of functors from  $\mathbb{T}$  to  $S$ , denote by  $K\text{-mod}$  the full subcategory of  $S^{\mathbb{T}}$  determined by  $K$ -modules.

Definition 4. A convex set is a finite product preserving functor from  $\mathbb{A}$  to  $S$ .

We denote the full subcategory of convex sets of  $S^{\mathbb{A}}$  by  $\mathcal{C}$ . Then the morphisms in  $\mathcal{C}$  are just the  $V$ -affine linear maps, that is,  $f: C \rightarrow C'$  in  $\mathcal{C}$ ,  $f$  satisfies  $f(\sum_{i=1}^n r_i c_i) = \sum_{i=1}^n r_i f(c_i)$ , where  $c_i \in C$ ,  $r_i \in K$ ,  $i = 1, \dots, n$  and  $\sum_{i=1}^n r_i = 1$ .

Remark. a). Note that, not every  $K$ -module  $M$  could be embedded into a  $R$ -module ( $R$  is the ring of real numbers), for example,  $M$  could be the  $K$ -module generated by two elements  $l$  (light) and  $h$  (heavy), such that  $r_l + r_h = r'h$ , for all  $r' \neq 0$ .

b). Both  $K\text{-mod}$  and  $\mathcal{C}$  are complete, since  $S^T$  and  $S^A$  are complete and limits always commute with finite products. For a functor  $F: \mathcal{J} \rightarrow \mathcal{C}$ , we take  $\varinjlim F$  in  $S^A$ , because limits commute with finite products, therefore  $\varinjlim F$  preserves finite products, so  $\varinjlim F \in \mathcal{C}$ .

Proposition 1. Any convex set  $C$  can be embedded into a  $K$ -module.

Proof. This is just a special case of the more general theorem in [8].

Definition 5. Let  $C, B$  and  $D$  be convex sets. A map  $f: C \times B \rightarrow D$  is called biconvex, if for any  $c, c_i \in C$ ,  $t_i \in K$ ,  $i = 1, \dots, n$ ,  $\sum_{i=1}^n t_i = 1$ ;  $b, b_j \in B$ ,  $s_j \in K$ ,  $j = 1, \dots, m$ ,  $\sum_{j=1}^m s_j = 1$ ,  $f(\sum_{i=1}^n t_i c_i, b) = \sum_{i=1}^n t_i f(c_i, b)$  and  $f(c, \sum_{j=1}^m s_j b_j) = \sum_{j=1}^m s_j f(c, b_j)$ .

Definition 6.  $C, B$  are two convex sets. The tensor product of  $C$  and  $B$  denoted by  $C \otimes B$  is an object in  $\mathcal{C}$ , together with  $i: C \times B \rightarrow C \otimes B$ , defined by the following universal property: for any convex set  $D$ , and any biconvex map  $f: C \times B \rightarrow D$ , there is a unique affine linear map  $\bar{f}: C \otimes B \rightarrow D$ , such that  $\bar{f} \circ i = f$ .

Theorem. For any pair of convex sets  $C$  and  $B$ ,  $C \otimes B$  exists.

Proof. Let  $F(C \times B)$  be the free convex set generated by  $C \times B$ . Define  $\sim$  as the congruence relation on  $F(C \times B)$  generated by

$$\left(\sum_{i=1}^n t_i c_i, b\right) \sim \sum_{i=1}^n t_i (c_i, b); \quad \left(c, \sum_{j=1}^m s_j b_j\right) \sim \sum_{j=1}^m s_j (c, b_j)$$

where  $c, c_i \in C$ ,  $t_i \in K$ ,  $i = 1, \dots, n$ ,  $\sum_{i=1}^n t_i = 1$ ;  $b, b_j \in B$ ,  $s_j \in K$ ,  $j = 1, \dots, m$

$\sum_{j=1}^m s_j = 1$ . Then  $F(C \times B) / \sim$  is a convex set. Let us denote the

congruence class of  $(c, b)$  by  $c \otimes b$ .

Claim:  $F(C \times B)/\sim$  is the tensor product of  $C$  and  $B$ .

First, we have the canonical biconvex map  $i: C \times B \longrightarrow F(C \times B)/\sim$

$i(c, b) = c \otimes b$ , for  $c \in C$  and  $b \in B$ .

If  $f: C \times B \longrightarrow D$  is a biconvex map, because  $F(C \times B)$  is a free convex set, so  $f$  induces an affine linear map

$\hat{f}: F(C \times B) \longrightarrow D$ . Since  $f$  is biconvex,

so  $\hat{f}$  sends the elements in the same congruence class of  $F(C \times B)$  to the same element in  $D$ , hence  $\hat{f}$  induces an affine linear map

$\bar{f}: F(C \times B)/\sim \longrightarrow D$ , and  $\bar{f} \circ i = f$ .

If  $g$  is another affine linear map from  $F(C \times B)/\sim$  to  $D$ , such that  $g \circ i = f$ , then for any  $c \in C$ ,  $b \in B$ ,

$$g(c \otimes b) = g \circ i(c, b) = f(c, b) = \bar{f} \circ i(c, b) = \bar{f}(c \otimes b).$$

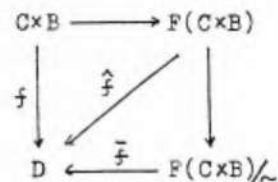
This shows that  $\bar{f}$  is unique.

Proposition 2. If  $f: C \longrightarrow C'$ ,  $g: B \longrightarrow B'$  are two affine linear maps between convex sets, define  $f \otimes g$  on the generators  $c \otimes b$  of  $C \otimes B$  as  $f \otimes g(c \otimes b) = f(c) \otimes g(b)$ , for  $c \in C$ ,  $b \in B$ . Then  $f \otimes g$  uniquely extends to an affine linear map from  $C \otimes B$  to  $C' \otimes B'$ .

Proof. This is true because that  $\{c \otimes b \mid c \in C, b \in B\}$  are generators of  $C \otimes B$  and  $f, g$  are affine linear.

Proposition 3. Define  $\otimes: \mathcal{C} \times \mathcal{C} \longrightarrow \mathcal{C}$ , for objects,  $(C, B) \longmapsto C \otimes B$ ; morphisms,  $(f, g) \longmapsto f \otimes g$ , where  $f: C \longrightarrow C'$ ,  $g: B \longrightarrow B'$ . Then  $\otimes$  is a bifunctor.

Proof. We have to show that if  $f: C \longrightarrow C'$ ,  $f': C' \longrightarrow C''$  and  $g: B \longrightarrow B'$ ,  $g': B' \longrightarrow B''$  are affine linear maps, then  $(f' \circ f) \otimes (g' \circ g) = (f' \otimes g') \circ (f \otimes g)$ .



Let us define a map  $h: C \times B \longrightarrow C \otimes B$  by  $h(c, b) = f' \cdot f(c) \otimes g' \cdot g(b)$ , then  $h$  is biconvex. For the canonical biconvex map  $i: C \times B \longrightarrow C \otimes B$ , it is easy to check that  $(f' \cdot f) \otimes (g' \cdot g) \cdot i = h = (f' \otimes g') \cdot (f \otimes g) \cdot i$ . By the universal property of the tensor product  $(C \otimes B, i)$ , we have  $f' \cdot f \otimes g' \cdot g = (f' \otimes g') \cdot (f \otimes g)$ .

Proposition 4. For each pair of convex sets  $C$  and  $B$ , the set of affine linear maps from  $C$  to  $B$ , denoted by  $\mathcal{L}(C, B)$ , is a convex set.

Proof. For  $f_i \in \mathcal{L}(C, B)$ ,  $t_i \in K$ ,  $i=1, \dots, n$ ,  $c \in C$ ,  $\sum_{i=1}^n t_i = 1$ , define

$$\left(\sum_{i=1}^n t_i f_i\right)(c) = \sum_{i=1}^n t_i f_i(c).$$

Then for  $c_j \in C$ ,  $s_j \in K$ ,  $j=1, \dots, m$ ,  $\sum_{j=1}^m s_j = 1$ ,

$$\begin{aligned} \left(\sum_{i=1}^n t_i f_i\right)\left(\sum_{j=1}^m s_j c_j\right) &= \sum_{i=1}^n t_i f_i\left(\sum_{j=1}^m s_j c_j\right) = \sum_{i=1}^n t_i \sum_{j=1}^m s_j f_i(c_j) \\ &= \sum_{j=1}^m s_j \left(\sum_{i=1}^n t_i f_i(c_j)\right) = \sum_{j=1}^m s_j \left(\sum_{i=1}^n t_i f_i\right)(c_j). \end{aligned}$$

This shows that  $\sum_{i=1}^n t_i f_i$  is an affine linear map.

Proposition 5. Let  $C, B$  and  $D$  be convex sets, then we have the natural isomorphism  $\varphi_{CBD}: \mathcal{L}(C \otimes B, D) \approx \mathcal{L}(C, \mathcal{L}(B, D))$ .

Proof. For an affine linear map  $f: C \otimes B \longrightarrow D$ , define  $\varphi_{CBD}(f) = \bar{f}: C \longrightarrow \mathcal{L}(B, D)$ ,  $\bar{f}(c)(b) = f(c \otimes b)$ , where  $c \in C$  and  $b \in B$ .  
1).  $\bar{f}(c)$  is affine linear for each  $c \in C$ .

Let  $b_i \in B$ ,  $t_i \in K$ ,  $i=1, \dots, n$ ,  $\sum_{i=1}^n t_i = 1$ ,  
then 
$$\begin{aligned} \bar{f}(c)\left(\sum_{i=1}^n t_i b_i\right) &= f\left(c \otimes \sum_{i=1}^n t_i b_i\right) = f\left(\sum_{i=1}^n t_i c \otimes b_i\right) \\ &= \sum_{i=1}^n t_i f(c \otimes b_i) = \sum_{i=1}^n t_i \bar{f}(c)(b_i). \end{aligned}$$

2).  $\bar{f}$  is affine linear.

Let  $\sum_{i=1}^n t_i c_i \in C$  and  $b \in B$ ,  
 then 
$$\begin{aligned} \bar{f}\left(\sum_{i=1}^n t_i c_i\right)(b) &= f\left(\left(\sum_{i=1}^n t_i c_i\right) \otimes b\right) = f\left(\sum_{i=1}^n t_i (c_i \otimes b)\right) \\ &= \sum_{i=1}^n t_i f(c_i \otimes b) = \sum_{i=1}^n t_i \bar{f}(c_i)(b). \end{aligned}$$

3).  $\varphi_{C \otimes B}$  is 1-1.

If  $\bar{f}_1 = \bar{f}_2$ , i.e. for any  $c \in C$ ,  $b \in B$ ,  $\bar{f}_1(c)(b) = \bar{f}_2(c)(b)$ ,  
 but this is  $f_1(c \otimes b) = f_2(c \otimes b)$ , so  $f_1 = f_2$ .

4).  $\varphi_{C \otimes B}$  is surjective.

For an affine linear map  $g: C \rightarrow \mathcal{L}(B, D)$ , define  
 $\hat{g}: C \otimes B \rightarrow D$ ,  $\hat{g}\left(\sum_{i=1}^n t_i c_i \otimes b_i\right) = \sum_{i=1}^n t_i g(c_i)(b_i)$ .  $\hat{g}$  is affine  
 linear and  $\tilde{g}: C \rightarrow \mathcal{L}(B, D)$ ,  $\tilde{g}(c)(b) = \hat{g}(c \otimes b) = g(c)(b)$ , for  
 any  $c \in C$ ,  $b \in B$ . So  $\tilde{g} = g$ .

This shows that  $\varphi_{C \otimes B}$  is a bijection. By straight diagram chasing,  
 one can show the naturality of  $\varphi$  in  $C, B$  and  $D$ .

Theorem.  $\mathcal{L}$  is a closed category.

§2.  $\mathcal{C}$  is a coreflective subcategory of  $K\text{-mod}/K$

Let  $u_*$  be the underlying functor from  $K\text{-mod}$  to  $\mathcal{C}$ . Since  $u_*$  preserves limits, we would like to ask whether it has a left adjoint.

For each convex set  $C$ , if  $C = \emptyset$ , define  $F_0(C) = \{0\}$  the zero  $K$ -module; if  $C \neq \emptyset$ , define an equivalence relation on  $C \times K$ :  
 $c_1, c_2 \in C, k_1, k_2 \in K, (c_1, k_1) \sim (c_2, k_2)$  iff 1).  $c_1 = c_2$  and  $k_1 = k_2$   
 or 2).  $k_1 = k_2 = 0$ . Then define the following operations on  $C \times K / \sim$ :

$$\text{for } r, k \in K, c \in C, r(c, k) = (c, rk),$$

$$\text{for } k_1, k_2 \in K, k_1 \neq 0 \text{ or } k_2 \neq 0, \text{ and } c_1, c_2 \in C$$

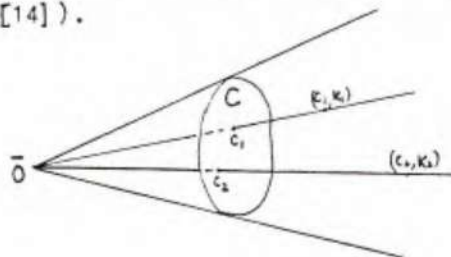
$$(c_1, k_1) + (c_2, k_2) = \left( \frac{k_1 c_1 + k_2 c_2}{k_1 + k_2}, k_1 + k_2 \right)$$

$$\bar{0} + \bar{0} = \bar{0}, \text{ where } \bar{0} \text{ is the equivalence class of } (c, 0).$$

These operations are well defined and it is easy to check that with these operations  $C \times K / \sim$  is a  $K$ -module, we denote this  $K$ -module by  $F_0(C)$ .

If  $f: C \rightarrow B$  is an affine linear map, then define  $\bar{f}: C \times K / \sim \rightarrow B \times K / \sim$  by  $\bar{f}(c, k) = (f(c), k)$ . It is easy to check that  $\bar{f}$  is a  $K$ -mod homomorphism. In this way we obtain a functor  $F_0$  from  $\mathcal{C}$  to  $K\text{-mod}$ .

Remark.  $C \times K / \sim$  is the cone generated by  $C$  as in the following picture (see [14]).



Proposition 1.  $F_0$  is left adjoint to  $U_0$ .

Proof. If convex set  $C \neq \emptyset$ , define  $i_C: C \rightarrow U_0 F_0(C) = C \times K/\sim$  by  $i_C(c) = (c, 1)$ , where  $c \in C$ ; if  $C = \emptyset$ , define  $i_C$  to be the unique affine linear map from empty convex set  $\emptyset$  to the one point convex set  $U_0 F_0(\emptyset)$ . Then it is easy to check that  $i$  is a natural transformation from  $F_0$  to  $U_0 F_0$ .

For any  $K$ -module  $M$  and any affine linear map  $f: C \rightarrow M$ , we have to construct a  $K$ -module homomorphism  $\bar{f}: F_0(C) \rightarrow M$ , such that  $U_0(\bar{f}) \circ i_C = f$ , i.e. the following diagram commutes.

$$\begin{array}{ccc} C & \xrightarrow{i_C} & U_0 F_0(C) \\ & \searrow f & \downarrow U_0 \bar{f} \\ & & U_0(M) \end{array}$$

Define  $\bar{f}: C \times K/\sim \rightarrow M$  by  $\bar{f}(c, k) = kf(c)$ .

Then  $\bar{f}(r(c, k)) = \bar{f}(c, rk) = r kf(c) = r(kf(c)) = r\bar{f}(c, k)$ ;

$$\bar{f}((c_1, k_1) + (c_2, k_2)) = \bar{f}\left(\frac{k_1 c_1 + k_2 c_2}{k_1 + k_2}, k_1 + k_2\right)$$

$$= (k_1 + k_2) f\left(\frac{k_1 c_1 + k_2 c_2}{k_1 + k_2}\right)$$

$$= (k_1 + k_2) f\left(\frac{k_1}{k_1 + k_2} c_1 + \frac{k_2}{k_1 + k_2} c_2\right)$$

$$= (k_1 + k_2) \left(\frac{k_1}{k_1 + k_2} f(c_1) + \frac{k_2}{k_1 + k_2} f(c_2)\right) \quad (\text{since}$$

$f$  is affine linear)

$$\begin{aligned}
 &= k_1 f(c_1) + k_2 f(c_2) \\
 &= \bar{f}(c_1, k_1) + \bar{f}(c_2, k_2)
 \end{aligned}$$

So,  $\bar{f}$  is a  $K$ -module homomorphism.

For any  $c \in C$ ,

$$\bar{f} \cdot i_C(c) = \bar{f}(c, 1) = 1f(c) = f(c), \text{ this is } \bar{f} \cdot i_C = f.$$

If  $g$  is other  $K$ -module homomorphism from  $C \times K / \sim$  to  $M$  such that  $g \cdot i_C = f$ , then for any  $(c, k) \in C \times K / \sim$ ,

$$\begin{aligned}
 g(c, k) &= g(k(c, 1)) = kg(c, 1) = kg \cdot i_C(c) = kf(c) \\
 &= k\bar{f} \cdot i_C(c) = k\bar{f}(c, 1) = \bar{f}(c, k),
 \end{aligned}$$

therefore  $g = \bar{f}$ .

This completes the proof.

The  $K$ -linear functionals  $M \xrightarrow{h} K$  on  $K$ -modules are the objects of a category  $K\text{-mod}/K$  in which a morphism from  $\begin{array}{c} M \\ \downarrow h \\ K \end{array}$  to  $\begin{array}{c} M' \\ \downarrow g \\ K \end{array}$

is a  $K$ -mod homomorphism  $f: M \rightarrow M'$  such that the triangle

$$\begin{array}{ccc}
 M & \xrightarrow{f} & M' \\
 \downarrow h & & \downarrow g \\
 K & & K
 \end{array}$$

commutes, i.e.  $g \circ f = h$ . For each  $K$ -linear functional

$h: M \rightarrow K$ ,  $h^{-1}(1)$  is a convex set and each  $K\text{-mod}/K$  morphism

$$\begin{array}{ccc}
 M & \xrightarrow{f} & M' \\
 \downarrow h & & \downarrow g \\
 K & & K
 \end{array}$$

induces an affine linear map from  $h^{-1}(1)$  to  $g^{-1}(1)$  ( $g^{-1}(1) = \emptyset$  forces  $h^{-1}(1) = \emptyset$ ; if  $h^{-1}(1) = \emptyset$ , the

induced map is just the unique affine linear map from  $\emptyset$  to  $g^{-1}(1)$ ;

otherwise if  $m \in h^{-1}(1)$ , since  $g \circ f = h$ , we have  $g(f(m)) = h(m) = 1$ ,

that is,  $f(m)$  is in  $g^{-1}(1)$ . This gives us a functor

$$G: K\text{-mod}/K \rightarrow \mathcal{C}.$$

Theorem 1. The functor  $G$  has a left adjoint  $F: \mathcal{C} \rightarrow K\text{-mod}/K$ , given by  $F(C) = F_0(C) \xrightarrow{p_C} K$ , where  $p_C = F_0(C \rightarrow 1)$ .

This theorem is a corollary of the following Lemma.

Lemma. If  $\mathcal{B}, \mathcal{E}$  are two categories, and  $\mathcal{B}$  has pullbacks and if given two functors  $\mathcal{B} \xrightleftharpoons[G_0]{F_0} \mathcal{E}$ ,  $F_0$  is left adjoint to  $G_0$ ,

then for any object  $B$  in  $\mathcal{B}$ , we have two induced functors

$\mathcal{B}/B \xrightleftharpoons[G]{F} \mathcal{E}/F_0(B)$ , such that  $F$  is left adjoint to  $G$ .

Proof. Define  $F(\downarrow h) = \downarrow_{F_0(B)}^{F_0(X)}$ ; and for each  $\downarrow_{F_0(B)}^E$  in  $\mathcal{E}/F_0(B)$ ,

form the pullback

$$\begin{array}{ccc} p(E) & \xrightarrow{\pi} & G_0(E) \\ \downarrow \hat{g} & & \downarrow G_0(g) \\ B & \xrightarrow{\gamma_B} & G_0 F_0(B) \end{array} \quad (\gamma_B \text{ is the unit of } F_0, G_0)$$

and then take  $G(\downarrow_{F_0(B)}^E) = \downarrow_{F_0(B)}^{p(E)}$ . It is immediate that  $F$  and

$G$  are functors.

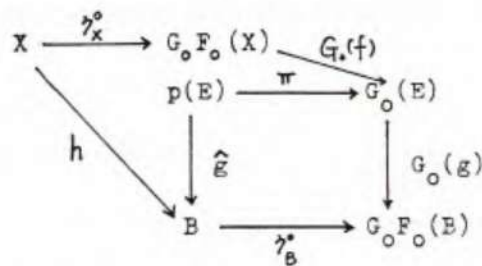
Now we construct a natural bijection

$$\alpha : \mathcal{E}/F_0(B) \left( \downarrow_{F_0(B)}^X, \downarrow_{F_0(B)}^E \right) \longrightarrow \mathcal{B}/B \left( \downarrow h, G(\downarrow_{F_0(B)}^E) \right).$$

Let  $f$  be a morphism from  $\downarrow_{F_0(B)}^X = \downarrow_{F_0(B)}^{F_0(X)}$  to  $\downarrow_{F_0(B)}^E$ , that is the

triangle  $\begin{array}{ccc} F_0(X) & \xrightarrow{f} & E \\ & \searrow F_0(h) & \swarrow g \\ & & F_0(B) \end{array}$  is commutative:  $g \cdot f = F_0(h)$ .

In the following diagram



we have  $G_0(g) \cdot G_0(f) \cdot \gamma_X^0 = G_0(g \cdot f) \cdot \gamma_X^0$  since  $G_0$  is a functor,  
 $= G_0 F_0(h) \cdot \gamma_X^0$  since  $g \cdot f = F_0(h)$ ,  
 $= \gamma_B^0 \cdot h$  since  $\gamma^0$  is natural.

Therefore, there is a unique  $\bar{f}: X \rightarrow p(E)$ , such that  $\hat{g} \cdot \bar{f} = h$ ,

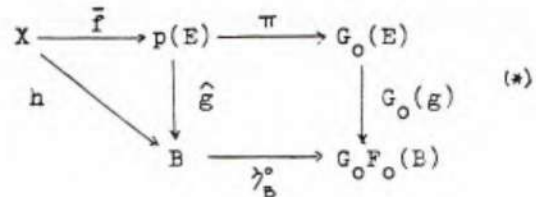
(and  $\pi \cdot \bar{f} = G_0(f) \cdot \gamma_X^0$ ). Thus  $\bar{f}$  is a morphism from

$$\begin{array}{c} X \\ \downarrow h \\ B \end{array} \text{ to } \begin{array}{c} p(E) \\ \downarrow \hat{g} \\ B \end{array} = G\left(\begin{array}{c} E \\ \downarrow g \\ F_0(B) \end{array}\right). \text{ We define } \alpha(f) = \bar{f}.$$

Now we construct  $\beta: \mathcal{B}/B\left(\begin{array}{c} X \\ \downarrow h \\ B \end{array}, G\left(\begin{array}{c} E \\ \downarrow g \\ F_0(B) \end{array}\right)\right) \rightarrow \mathcal{E}/F_0(B)\left(F\left(\begin{array}{c} X \\ \downarrow h \\ E \end{array}\right), \begin{array}{c} E \\ \downarrow g \\ F_0(B) \end{array}\right)$ .

Let  $\bar{f}$  be a morphism from  $\begin{array}{c} X \\ \downarrow h \\ B \end{array}$  to  $G\left(\begin{array}{c} E \\ \downarrow g \\ F_0(B) \end{array}\right) = \begin{array}{c} p(E) \\ \downarrow \hat{g} \\ B \end{array}$  in  $\mathcal{B}/B$ ,

then in the diagram



$\hat{g} \circ \bar{f} = h$ , and we define  $\beta(\bar{f}) = f$  as the transpose of the composite  $X \xrightarrow{\bar{f}} p(E) \xrightarrow{\pi} G_0(E)$ , that is  $f = \xi_E^0 \circ F_0(\pi \circ \bar{f})$ , where  $\xi^0$  is the counit of  $F_0, G_0$ .

We want to show that the triangle  $F_0(X) \xrightarrow{f} E$  is commutative, since:

$$\begin{array}{ccc} F_0(X) & \xrightarrow{f} & E \\ & \searrow F_0(h) & \swarrow g \\ & & F_0(B) \end{array}$$

commutative, since:

$$\begin{aligned} g \circ f &= g \circ \xi_E^0 \circ F_0(\pi \circ \bar{f}) && \text{(the definition of } f) \\ &= \xi_{F_0(B)}^0 \circ F_0(G_0(g) \circ F_0(\pi \circ \bar{f})) && \text{(since } \xi^0 \text{ is natural)} \\ &= \xi_{F_0(B)}^0 \circ F_0(G_0(g) \circ \pi \circ \bar{f}) && \text{(since } F_0 \text{ is a functor)} \\ &= \xi_{F_0(B)}^0 \circ F_0(\gamma_B^0 \circ h) && \text{(since } G_0(g) \circ \pi \circ \bar{f} = \gamma_B^0 \circ \hat{g} \circ \bar{f} = \gamma_B^0 \circ h, \\ & && \text{see the above diagram (*))} \\ &= \xi_{F_0(B)}^0 \circ F_0(\gamma_B^0) \circ F_0(h) && \text{(again since } F_0 \text{ is a functor)} \\ &= 1_{F_0(B)} \circ F_0(h) && \text{(by the equivalent condition} \\ & && \text{of the adjointness)} \\ &= F_0(h) \end{aligned}$$

This shows that  $f$  is a morphism from  $F(\downarrow h)$  to  $\downarrow_g$  in  $\mathcal{C}/F(B)$ .

The proofs of the naturality of  $\alpha$  and  $\beta$ , and that  $\beta$  is the inverse of  $\alpha$  are straight forward diagram chasing exercises. this completes the proof of the lemma.

Now the proof of the theorem:

Using the above Lemma, let  $\mathcal{C} = K\text{-mod}$ ,  $\mathcal{B} = \mathcal{C}$ ,  $G_0 = u$ , and  $F_0 = F_0$ . Take  $B = 1$ , the one element convex set, then  $F_0(B) = F_0(1) = K$ , the free  $K$ -module on one generator. We have  $F(C) = F_0(C \rightarrow 1) = F_0(C) \xrightarrow{p_C} K$ , where  $F_0(C)$  is the free  $K$ -module generated by the convex set  $C$ .

Since  $C \rightarrow 1$  sends every  $c$  in  $C$  to the unique generator of the free  $K$ -module  $K$ , we have that  $p_C = F_0(C \rightarrow 1): F_0(C) = C \times K / \sim \rightarrow K$  takes  $(c, k) = k(c, 1)$  to  $k1 = k$ .

For each  $\begin{array}{c} M \\ \downarrow h \\ K \end{array}$  in  $K\text{-mod}/K$ , the pullback diagram

$$\begin{array}{ccc} p(M) & \longrightarrow & G_0(M) \\ \downarrow & & \downarrow G(h) \\ 1 & \xrightarrow{\eta_1} & G_0(K) \end{array}$$

and the fact that  $G_0(K) = G_0 F_0(1)$  tell us that  $p(M) = h^{-1}(1)$ , that is  $G(\begin{array}{c} M \\ \downarrow h \\ K \end{array}) = h^{-1}(1)$ .

Therefore we have  $\mathcal{C} \xrightleftharpoons[G]{F} K\text{-mod}/K$ ,  $F$  is left adjoint to  $G$ ,

which completes the proof of the theorem.

Proposition 2. For any convex set  $C$ ,  $C \approx GF(C)$ .

Proof. If  $C \neq \emptyset$ ,  $GF(C) = G(\begin{array}{c} C \times K / \sim \\ \downarrow p_C \\ K \end{array}) = p_C^{-1}(1) = \{(c, 1) \mid c \in C\}$ .

The map  $C \rightarrow GF(C)$ ,  $c \mapsto (c, 1)$  is a bijection in  $\mathcal{C}$ .

if  $C = \emptyset$ ,  $GF(\emptyset) = G(\begin{array}{c} \{0\} \\ \downarrow 0 \\ K \end{array}) = \emptyset$ .

Corollary.  $\mathcal{C}$  is equivalent to a coreflective subcategory of  $K\text{-mod}/K$ .

Proposition 3. Let  $\downarrow h \in K\text{-mod}/K$ . Then  $FG(\downarrow h) \approx \downarrow h$

if and only if  $h^{-1}(0) = \{0\}$ .

Proof. Let  $FG(\downarrow h) \approx \downarrow h$  Since  $FG(\downarrow h) = \downarrow_{K}^{h^{-1}(1) \times K / \sim} p_{h^{-1}(0)}$ ,

so there is a  $K$ -module isomorphism  $f: h^{-1}(1) \times K / \sim \rightarrow M$ , such that  $p_{h^{-1}(0)} = h \circ f$  or  $p_{h^{-1}(0)} \circ f^{-1} = h$ . Since the kernel of  $p_{h^{-1}(0)}$  is  $\{\bar{0}\}$  and  $f^{-1}$  is an isomorphism, so the kernel of  $h$  is  $\{0\}$ .

Conversely, if  $h^{-1}(0) = \{0\}$ , we define  $g: M \rightarrow h^{-1}(1) \times K / \sim$  as follows:

$$g(m) = \left( \frac{m}{h(m)}, h(m) \right), \text{ for } m \in M, m \neq 0; g(0) = \bar{0}.$$

$g$  is a  $K$ -module homomorphism:

given  $r \in K$  and  $m, m_1, m_2 \in M$ , we have:

$$g(rm) = \left( \frac{rm}{h(rm)}, h(rm) \right) = \left( \frac{rm}{rh(m)}, rh(m) \right) = r \left( \frac{m}{h(m)}, h(m) \right) = rg(m).$$

$$\begin{aligned} g(m_1 + m_2) &= \left( \frac{m_1 + m_2}{h(m_1 + m_2)}, h(m_1 + m_2) \right) \\ &= \left( \frac{h(m_1) \frac{m_1}{h(m_1)} + h(m_2) \frac{m_2}{h(m_2)}}{h(m_1) + h(m_2)}, h(m_1) + h(m_2) \right) \\ &= \left( \frac{m_1}{h(m_1)}, h(m_1) \right) + \left( \frac{m_2}{h(m_2)}, h(m_2) \right) = g(m_1) + g(m_2) \end{aligned}$$

(since  $\frac{m_1}{h(m_1)}, \frac{m_2}{h(m_2)}$  are in  $h^{-1}(1)$ )

$g$  is surjective: for any  $(c,k) \in h^{-1}(1) \times K/\sim$ , let  $m = kc$ , then  $h(kc) = kh(c) = k$  (note that  $c \in h^{-1}(1)$ ). Therefore,

$$g(m) = \left( \frac{m}{h(m)}, h(m) \right) = \left( \frac{kc}{h(kc)}, h(kc) \right) = \left( \frac{kc}{kh(c)}, kh(c) \right) = (c,k).$$

$g$  is 1-1: if  $m \neq 0$ , then  $h(m) \neq 0$ , so  $g(m) = \left( \frac{m}{h(m)}, h(m) \right) \neq \bar{0}$ .

Also we have, for any  $m \in M$ ,  $p_{h(m)} \circ g(m) = p_{h(m)} \left( \frac{m}{h(m)}, h(m) \right) = h(m)$ ,

i.e.  $p_{h(m)} \circ g = h$ .

Therefore,  $g$  is an isomorphism between  $FG \left( \begin{array}{c} M \\ \downarrow h \\ K \end{array} \right)$  and  $\begin{array}{c} M \\ \downarrow h \\ K \end{array}$ .

Definition. If  $M$  is a  $K$ -module and  $N$  is a  $K$ -submodule of  $M$ , the quotient  $K$ -module  $M/N$  and the canonical  $K$ -mod homomorphism  $q: M \rightarrow M/N$  is defined by the following universal property: for any  $K$ -module  $B$ , and  $K$ -mod homomorphism  $f: M \rightarrow B$ , such that  $f(n) = 0$  for all  $n$  in  $N$ , there is a unique  $K$ -module homomorphism  $\bar{f}: M/N \rightarrow B$  such that  $\bar{f} \circ q = f$ .

$$\begin{array}{ccc} M & \xrightarrow{q} & M/N \\ & \searrow f & \swarrow \bar{f} \\ & & B \end{array}$$

Remark: We can construct  $M/N$  in the usual way (construction of quotient  $R$ -module, where  $R$  is a ring), namely define  $M/N$  as the set of congruence classes determined by  $N$  on  $M$ , except that since  $K$  has no negatives, the congruence relation on  $M$  should be defined as follows:

$m_1 \sim m_2$  iff there exist  $n_1, n_2 \in N$ , such that  $m_1 + n_1 = m_2 + n_2$ .

**Theorem 2.** The functor  $F: \mathcal{C} \longrightarrow K\text{-mod}/K$  also has a left adjoint  $L: K\text{-mod}/K \longrightarrow \mathcal{C}$ .

**Proof.** Let  $\begin{array}{c} M \\ \downarrow h \\ K \end{array}$  be an object in  $K\text{-mod}/K$ . Form the quotient  $K$ -module  $M/h^{-1}(0)$ , then  $h$  induces a  $K$ -module homomorphism

$$\bar{h}: M/h^{-1}(0) \longrightarrow K, \text{ such that the triangle } \begin{array}{ccc} M & \longrightarrow & M/h^{-1}(0) \\ & \searrow h & \swarrow \bar{h} \\ & & K \end{array}$$

is commutative.

Let us denote the congruence class of  $m$  in  $M/h^{-1}(0)$  by  $\bar{m}$ .

If  $g$  is a morphism from  $\begin{array}{c} M \\ \downarrow h \\ K \end{array}$  to  $\begin{array}{c} M' \\ \downarrow h' \\ K \end{array}$  (that is  $h' \circ g = h$ ), then

$g$  induces a homomorphism  $\tilde{g}: M/h^{-1}(0) \longrightarrow M'/h'^{-1}(0)$  given by  $\tilde{g}(\bar{m}) = \overline{g(m)}$  (For  $h(m) = 0$  implies  $h' \circ g(m) = 0$ ).

We define  $L(\begin{array}{c} M \\ \downarrow h \\ K \end{array}) = \bar{h}^{-1}(1)$  and  $L(g) =$  the restriction of  $\tilde{g}$  on  $\bar{h}^{-1}(1)$ ,  $\tilde{g}$  is an affine linear map between the convex sets  $\bar{h}^{-1}(1)$  and  $\bar{h}'^{-1}(1)$ . It is easy to check that  $L$  is a functor.

Now we construct a natural bijection

$$\alpha : \mathcal{C}(L(\begin{array}{c} M \\ \downarrow h \\ K \end{array}), C) \xrightarrow{(\hat{\quad})} K\text{-mod}/K(\begin{array}{c} M \\ \downarrow h \\ K \end{array}, F(C)).$$

If we have a morphism  $f: L(\begin{array}{c} M \\ \downarrow h \\ K \end{array}) = \bar{h}^{-1}(1) \longrightarrow C$  in  $\mathcal{C}$ ,

define  $\hat{f} = \alpha(f): M \longrightarrow F(C) = C \times K / \sim$  as follows,

- 1).  $h(m) = 0, \hat{f}(m) = \bar{0}$ ;
- 2).  $h(m) \neq 0, \hat{f}(m) = (f(\frac{\bar{m}}{h(m)}), h(m))$ , note that  $\bar{h}(\bar{m}) = h(m)$ .

Let's check that  $f$  is a  $K$ -mod homomorphism:

1).  $\hat{f}$  preserves scalar multiplication. Let  $r \in K$ ,  $m \in M$ ,

(a).  $h(m) = 0$ , then  $h(rm) = 0$ , so  $\hat{f}(rm) = \bar{0} = r\bar{0} = r\hat{f}(m)$ ;

$$\begin{aligned} \text{(b). } h(m) \neq 0, r \neq 0, \text{ then } \hat{f}(rm) &= (f(\frac{\overline{rm}}{h(rm)}), h(rm)) \\ &= (f(\frac{r\bar{m}}{rh(m)}), rh(m)) \\ &= (f(\frac{\bar{m}}{h(m)}), rh(m)) \\ &= r(f(\frac{\bar{m}}{h(m)}), h(m)) \\ &= r\hat{f}(m) \end{aligned}$$

$$r = 0, \text{ then } \hat{f}(0m) = \hat{f}(0) = \bar{0} = 0\hat{f}(m).$$

2).  $\hat{f}$  preserves addition. Let  $m_1, m_2 \in M$ ,

(a). both  $h(m_1)$  and  $h(m_2)$  are 0,

then  $h(m_1+m_2) = 0$ , so  $\hat{f}(m_1+m_2) = \bar{0} = \hat{f}(m_1) + \hat{f}(m_2)$ .

(b).  $h(m_1) = 0$ , but  $h(m_2) \neq 0$ , then  $h(m_1+m_2) = h(m_2) \neq 0$ ,

$$\begin{aligned} \hat{f}(m_1+m_2) &= (f(\frac{\overline{m_1+m_2}}{h(m_1+m_2)}), h(m_1+m_2)) \\ &= (f(\frac{\bar{m}_1+\bar{m}_2}{h(m_1)+h(m_2)}), h(m_1)+h(m_2)) \\ &= (f(\frac{\bar{m}_2}{h(m_2)}), h(m_2)) \quad (\text{since } \bar{m}_1 = 0, h(m_1) = 0) \\ &= \hat{f}(m_2) = \hat{f}(m_1) + \hat{f}(m_2) \quad (\hat{f}(m_1) = 0) \end{aligned}$$

(c).  $h(m_1) \neq 0$  and  $h(m_2) \neq 0$ ,

$$\begin{aligned}
\hat{f}(m_1+m_2) &= (f(\frac{\bar{m}_1+\bar{m}_2}{h(m_1+m_2)}), h(m_1+m_2)) \\
&= (f(\frac{\bar{m}_1+\bar{m}_2}{h(m_1)+h(m_2)}), h(m_1)+h(m_2)) \\
&= (f(\frac{h(m_1)\bar{m}_1}{h(m_1)+h(m_2)} + \frac{h(m_2)\bar{m}_2}{h(m_1)+h(m_2)}), h(m_1)+h(m_2)) \\
&= (\frac{h(m_1)}{h(m_1)+h(m_2)}f(\frac{\bar{m}_1}{h(m_1)}) + \frac{h(m_2)}{h(m_1)+h(m_2)}f(\frac{\bar{m}_2}{h(m_2)}), h(m_1)+h(m_2)) \\
&= (f(\frac{\bar{m}_1}{h(m_1)}), h(m_1)) + (f(\frac{\bar{m}_2}{h(m_2)}), h(m_2)) \\
&= \hat{f}(m_1) + \hat{f}(m_2) .
\end{aligned}$$

Finally, let's see that  $\hat{f}$  is a morphism from  $\begin{array}{c} M \\ \downarrow h \\ K \end{array}$  to  $\begin{array}{c} C \times K / \sim \\ \downarrow \\ K \end{array}$ ,

that is that  $p_C \circ \hat{f} = h$ . For any  $m \in M$ ,

$$p_C \circ \hat{f}(m) = p_C(f(\frac{\bar{m}}{h(m)}), h(m)) = h(m), \text{ this shows that } p_C \circ \hat{f} = h.$$

On the other hand, we constructe

$$\beta : K\text{-mod}/K(\begin{array}{c} M \\ \downarrow h \\ K \end{array}, F(C)) \xrightarrow{(\checkmark)} \mathcal{C}(\begin{array}{c} M \\ \downarrow h \\ K \end{array}, C) .$$

Let  $g$  be a morphism from  $\begin{array}{c} M \\ \downarrow h \\ K \end{array}$  to  $F(C) = \begin{array}{c} C \times K / \sim \\ \downarrow \\ K \end{array}$  (i.e.

$p_C \circ g = h$ ), and note that the elements in  $\bar{h}^{-1}(1)$  have the form

$\frac{\bar{m}}{h(\bar{m})} = \frac{\bar{m}}{h(m)}$ . Let  $\pi_C$  be the projection from  $C \times K/\mathcal{L}$  to  $C$ ,  
 i.e.  $\pi_C(c, k) = c$ . We define  $\beta(g) = \check{g}: h^{-1}(1) \rightarrow C$  by  
 $\check{g}(\frac{\bar{m}}{h(\bar{m})}) = \pi_C \circ g(\frac{\bar{m}}{h(\bar{m})})$ .

If  $\bar{m} = \bar{m}_1$  in  $M/h^{-1}(0)$ , then there exist  $u, u_1 \in h^{-1}(0)$ ,  
 $m+u = m_1+u_1$ . Since  $p_C \circ g = h$ ,  $g(u) = g(u_1) = 0$ , so  $g(m) =$   
 $g(m_1)$ , this gives us that  $\check{g}(\frac{\bar{m}}{h(\bar{m})}) = \check{g}(\frac{\bar{m}}{h(\bar{m})})$  (note that  
 $h(m) = h(m_1)$ ), i.e.  $\check{g}$  is well defined.

The naturality of  $\alpha$  and  $\beta$ , and the fact that  $\beta$  is the  
 inverse of  $\alpha$  are straight forward exercises.

§ 3. A closed functor  $\varphi$  from  $\mathcal{C}$  to  $\mathcal{U}$ -cat.

Let  $\mathcal{U}$  be the poset  $([0, \infty], \geq)$ .  $\mathcal{U}$  has a symmetric tensor defined by  $r_1 \otimes r_2 = r_1 + r_2$ , any  $r_1, r_2 \in \mathcal{U}$ ; and Hom defined

$$\text{by } [r_1, r_2] = \begin{cases} r_2 - r_1, & \text{if } r_2 \geq r_1. \\ 0, & \text{if } r_2 < r_1 \end{cases}$$

It is easy to check that  $r_1 + r_2 \leq r_3$  iff  $r_1 \leq r_2, r_3$ . Thus,  $([0, \infty], \geq)$  is a closed category.

We denote the category of  $\mathcal{U}$ -categories and  $\mathcal{U}$ -functors (see [11]) by  $\mathcal{U}$ -cat.

If  $A \in \mathcal{U}$ -cat, that is for any objects  $a, b$  in  $A$ ,  $A(a, b) \in [0, \infty]$ ,

it satisfies (1).  $A(a, b) + A(b, c) \geq A(a, c)$

(2).  $0 \geq A(a, a)$ .

If  $f: A \rightarrow B$  is a  $\mathcal{U}$ -functor between two  $\mathcal{U}$ -categories, that is  $A(a, b) \geq B(f(a), f(b))$ , for all  $a, b \in A$ . Thus  $\mathcal{U}$ -cat is the category of generalized metric spaces and Lip 1 functions.

Proposition 1.  $\mathcal{U}$ -cat is a closed category with the tensor product  $\otimes$  and Hom as following (see [11]):

Let  $A, B$  be generalized metric spaces and  $f, g$  be Lip 1 functions from  $A$  to  $B$ , then  $A \otimes B$  is again a generalized metric space or a  $\mathcal{U}$ -category, with objects:  $\text{ob}(A \otimes B) = \text{ob}(A) \times \text{ob}(B)$ , where  $\text{ob}(A)$  is the set of objects of  $A$ , and

$$A \otimes B((a_1, b_1), (a_2, b_2)) = A(a_1, a_2) + B(b_1, b_2);$$

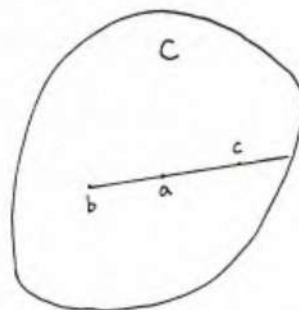
$\text{Hom}(A, B)$  is also a generalized metric space or a  $\mathcal{U}$ -category

with objects:  $\text{ob}(\text{Hom}(A, B)) = \{\text{Lip 1 functions from } A \text{ to } B\}$

and  $\text{Hom}(A, B)(f, g) = \sup_{a \in A} B(f(a), g(a))$ .

Definition 1.  $C$  is a convex set,  $a, b \in C$ . Define

$$\text{dist}(a, b) = -\log \sup_{\substack{t \\ \exists c \in C, \exists t \in [0, 1] \\ a = tb + (1-t)c}}$$



Remark. If  $C$  is a convex subset of  $\mathbb{R}^2$ , then in the above definition

$$t = \frac{d(a, c)}{d(b, c)}, \text{ where } d(a, c) \text{ is the}$$

Euclidean distance between  $a$  and  $c$ ,  $\text{dist}(a, b)$  tells us the situation of the line segment or ray starting from  $b$  passing through  $a$  in  $C$ .

Proposition 2. For any convex set  $C$ ,  $\text{dist}$  is a generalized metric on  $C$ .

Proof. It is clear that  $\text{dist}(c, c) = 0$ .

For any  $a, b, c \in C$ , we have to show that the triangle inequality is true, i.e.  $\text{dist}(a, b) + \text{dist}(b, c) \geq \text{dist}(a, c)$ .

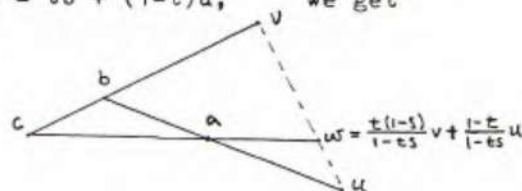
$$\text{We denote } \sup_{\substack{t \\ \exists t \in [0, 1], \exists u \in C \\ a = tb + (1-t)u}} = t_0 ;$$

$$\sup_{\substack{s \\ \exists s \in [0, 1], \exists v \in C \\ b = sc + (1-s)v}} = s_0 ;$$

$$\sup_{\substack{r \\ \exists r \in [0, 1], \exists w \in C \\ a = rc + (1-r)w}} = r_0 .$$

Substitute  $b = sc + (1-s)v$  into  $a = tb + (1-t)u$ , we get

$$\begin{aligned} a &= t(sc + (1-s)v) + (1-t)u \\ &= tsc + t(1-s)v + (1-t)u \\ &= tsc + (1-ts)\left(\frac{t(1-s)}{1-ts}v + \frac{1-t}{1-ts}u\right) \end{aligned}$$



since  $t(1-s) = t-ts < 1-ts$ ,  $1-t < 1-ts$  and  $\frac{t(1-s)}{1-ts} + \frac{1-t}{1-ts} =$   
 $= \frac{t-ts+1-t}{1-ts} = 1$ , so  $\frac{t(1-s)}{1-ts}v + \frac{1-t}{1-ts}u \in C$ , denote it by  $w$ .

Thus,  $ts \leq r_0$ . Taking sup over  $t, s$ , we get  $t_0 s_0 \leq r_0$ .

Therefore,  $-\log t_0 s_0 = (-\log t_0) + (-\log s_0) \geq -\log r_0$ ,

i.e.  $\text{dist}(a,b) + \text{dist}(b,c) \geq \text{dist}(a,c)$ .

Lemma 1.  $C$  is a convex set,  $a, b \in C$ , then

$$\text{dist}(a,b) = \begin{matrix} -\log \sup t \\ \exists t \in [0,1], \exists c \in C \\ a=tb+(1-t)c \end{matrix} = \begin{matrix} \inf(-\log t) \\ \exists t \in [0,1], \exists c \in C \\ a=tb+(1-t)c \end{matrix} .$$

Lemma 2. If  $f: C \rightarrow B$  is an affine linear map, then for any  $a, b \in C$ ,  $\text{dist}(a,b) \geq \text{dist}(f(a), f(b))$ , i.e.  $f$  is a Lip 1 function with respect to  $\text{dist}$  on  $C$  and  $B$ .

Proof. If  $a = tb + (1-t)c$ , because  $f$  is affine linear, so  $f(a) = tf(b) + (1-t)f(c)$ , then

$$-\log t \geq \begin{matrix} \inf(-\log s) \\ \exists s \in [0,1], \exists u \in B \\ f(a)=sf(b)+(1-s)u \end{matrix}$$

$$\text{therefore, } \text{dist}(a,b) = \begin{matrix} \inf(-\log t) \\ \exists t \in [0,1], \exists c \in C \\ a=tb+(1-t)c \end{matrix} \geq \begin{matrix} \inf(-\log s) \\ \exists s \in [0,1], \exists u \in B \\ f(a)=tf(b)+(1-s)u \end{matrix} = \text{dist}(f(a), f(b)) .$$

With  $\text{dist}$  as metric on convex sets and using Lemma 2, we get a functor  $\varphi: \mathcal{C} \rightarrow \mathcal{U}\text{-cat}$ , which sends each convex set  $C$  to  $C = (C, \text{dist})$  the corresponding generalized metric space of  $C$ , and sends each affine linear map  $f$  to the corresponding Lip 1 function  $\varphi f$ .

Proposition 2.  $\varphi$  is a closed functor. (see [5] for the definition of closed functor)

Proof. Since the unit of tensor in  $\mathcal{U}$ -cat is one element metric space 1 and the unit of tensor in  $\mathcal{C}$  is one element convex set 1, so we have

$$\varphi^{\circ}: 1 \longrightarrow \varphi 1 = 1 .$$

For each pair of convex sets  $C$  and  $B$ , we have to construct a Lip 1 function ( $\mathcal{U}$ -functor)

$$\varphi_{CB}: \varphi \mathcal{C}(C, B) \longrightarrow \text{Hom}(\varphi C, \varphi B),$$

where  $\mathcal{C}(C, B)$  is the convex set of affine linear maps from  $C$  to  $B$ ,  $\varphi \mathcal{C}(C, B)$  means to take it as a metric space with  $\text{dist}$  as metric.  $\text{Hom}(\varphi C, \varphi B)$  is the metric space of Lip 1 functions from metric space  $\varphi C$  to metric space  $\varphi B$  with sup metric. From Lemma 2, we have an obvious map  $\varphi_{CB}$ , sends  $f$  to  $\varphi f$ , we have to prove that it is Lip 1.

Let  $f, g \in \varphi \mathcal{C}(C, B)$ ,

$$\text{dist}(f, g) = \inf_{\substack{t \in [0, 1], h \in \varphi \mathcal{C}(C, B) \\ f = tg + (1-t)h}} (-\log t)$$

$$\begin{aligned} \text{and } \text{Hom}(\varphi C, \varphi B)(\varphi f, \varphi g) &= \sup_{c \in C} B(f(c), g(c)) \\ &= \sup_{c \in C} \text{dist}(f(c), g(c)) \end{aligned}$$

Because  $f = tg + (1-t)h$  iff  $f(c) = tg(c) + (1-t)h(c)$ , for all  $c \in C$ ,

$$\text{so } \text{dist}(f, g) = \inf_{\substack{t \in [0, 1], h \in \varphi \mathcal{C}(C, B) \\ f = tg + (1-t)h}} (-\log t) \geq \inf_{\substack{s \in [0, 1], b \in B \\ f(c) = sg(c) + (1-s)b}} (-\log s) = \text{dist}(f(c), g(c)),$$

$$\text{thus } \text{dist}(f, g) \geq \sup_{c \in C} \text{dist}(f(c), g(c)) = \text{Hom}(\varphi C, \varphi B)(\varphi f, \varphi g) .$$

From straight diagram chasing, it is easy to see that  $\varphi_{CB}$  is natural in  $C$  and  $B$ , and also the following three diagrams are commutative.

$$\begin{array}{ccc} 1 & \xrightarrow{j_C} & \varphi_{\mathcal{C}}(C, C) \\ \downarrow & & \downarrow \varphi_{CC} \\ 1 & \xrightarrow{j_C} & \text{Hom}(\varphi_C, \varphi_C) \end{array}$$

where  $j_C$  takes  $1$  to  $1_C$ , the identity map of  $C$ ;

$$\begin{array}{ccc} \varphi_{\mathcal{C}}(1, C) & \xrightarrow{\varphi_{1C}} & \text{Hom}(\varphi_1, \varphi_C) \\ \varphi_{i_C} \uparrow & & \downarrow \text{Hom}(\varphi^{\circ}, 1\varphi_C) \\ \varphi_C & \xrightarrow{i_C} & \text{Hom}(1, \varphi_C) \end{array}$$

where  $i_C: C \rightarrow \mathcal{C}(1, C)$  is an isomorphism;

$$\begin{array}{ccccc} \varphi_{\mathcal{C}}(A, B) & \xrightarrow{\varphi_{\mathcal{C}}(C, -)_{AB}} & \varphi_{\mathcal{C}}(\mathcal{C}(C, A), \mathcal{C}(C, B)) & \xrightarrow{\varphi_{\mathcal{C}}(C, A)\mathcal{C}(C, B)} & \text{Hom}(\varphi_{\mathcal{C}}(C, A), \varphi_{\mathcal{C}}(C, B)) \\ \downarrow \varphi_{AB} & & & & \downarrow \text{Hom}(1, \varphi_{CB}) \\ \text{Hom}(\varphi_A, \varphi_B) & \xrightarrow{\text{Hom}(\varphi_C, -)_{\varphi_A \varphi_B}} & \text{Hom}(\text{Hom}(\varphi_C, \varphi_A), \text{Hom}(\varphi_C, \varphi_B)) & \xrightarrow{\text{Hom}(\varphi_{CA}, 1)} & \text{Hom}(\varphi_{\mathcal{C}}(C, A), \text{Hom}(\varphi_C, \varphi_B)) \end{array}$$

where  $A, B$  and  $C$  are convex sets.

This completes the proof that  $\varphi$  is a closed functor.

4. Professor Schanuel proved that  $\varphi$  has a left adjoint.

Lemma 1. Let  $C$  be a convex set,  $a, b \in C$ ,  $T = \{t \in [0, 1] \mid \exists c \in C, a = tb + (1-t)c\}$ , then  $T$  is downward closed.

Proof. If  $a = tb + (1-t)c$ , and  $s < t$ , then

$$a = sb + (t-s)b + (1-t)c = sb + (1-s)\left(\frac{t-s}{1-s}b + \frac{1-t}{1-s}c\right), \text{ so } s \in T.$$

Lemma 2. Let  $a, b_1, b_2 \in C$ . If  $ta + (1-t)b_1 = ta + (1-t)b_2$  holds for some  $t \in (0, 1)$ , then it holds for all  $t \in (0, 1)$ . (but possibly not at  $t = 0$ )

Proof. Let  $\Delta$  be an equilateral triangle generated by  $x, y_1, y_2$ .  $f: \Delta \rightarrow C$  is an affine linear map, such that  $f(x) = a$ ,  $f(y_i) = b_i$ ,  $i=1, 2$ . Then the hypothesis is

$$f(tx + (1-t)y_1) = f(tx + (1-t)y_2), \text{ let}$$

$$u_1 = tx + (1-t)y_1, u_2 = tx + (1-t)y_2,$$

$$\text{that is } f(u_1) = f(u_2).$$

For any vertical line segment on the left of  $u_1u_2$ ,  $f$  is constant on  $v_1v_2$ , since  $v_1$  is a convex

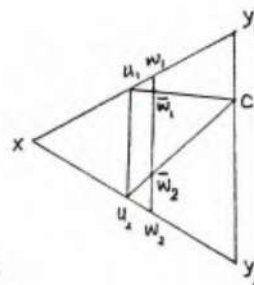
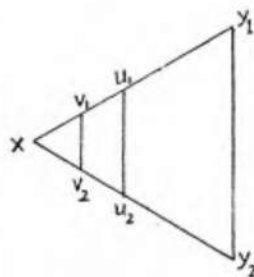
combination of  $x$  and  $u_1$  and  $f(u_1) = f(u_2)$ , so  $f(v_1) = f(v_2)$ .

Same is also true for  $u_1u_2$  itself.

Now we have to show that  $f$  is also constant on any line segment in between  $u_1u_2$  and  $y_1y_2$ .

We see that  $f$  is constant on  $\bar{w}_1\bar{w}_2$ ,

assume  $w_1 = sx + (1-s)y_1$ ,  $w_2 = sx + (1-s)y_2$



then  $\frac{\bar{w}_1 \bar{w}_2}{1-t} = \frac{s}{t}$ , (assume that the length of each side is 1)

i.e.  $\bar{w}_1 \bar{w}_2 = \frac{s}{t}(1-t) = \varepsilon$  does not depend on the position of

$\bar{w}_1, \bar{w}_2$ . This shows that  $f$  is constant on any line segment of  $w_1 w_2$  which has length  $\varepsilon$ . Thus,  $f$  is constant on  $w_1 w_2$ .

Theorem.  $\varphi: \mathcal{C} \rightarrow \mathcal{U}$ -cat preserves all limits.

Proof. First,  $\varphi$  preserves arbitrary products.

Let  $C_i \in \mathcal{C}$ , since the metric of the product of metric spaces is the sup metric, we have to show that if  $a, b \in \prod_i C_i$ ,  $a = (a_i)$ ,  $b = (b_i)$ ,  $\text{dist}(a, b) = \sup_i \text{dist}(a_i, b_i)$ .

$$\text{dist}(a, b) = -\log \sup_{\substack{t \in [0, 1], \exists c \in \prod_i C_i \\ a = tb + (1-t)c}} t ;$$

$$\begin{aligned} \sup_i \text{dist}(a_i, b_i) &= \sup_i (-\log \sup_{\substack{t_i \in [0, 1], \exists c_i \in C_i \\ a_i = t_i b_i + (1-t_i) c_i}} t_i) \\ &= -\log(\inf_i \sup_{\substack{t_i \in [0, 1], \exists c_i \in C_i \\ a_i = t_i b_i + (1-t_i) c_i}} t_i) \end{aligned}$$

i.e. we have to show  $\sup_i T = \inf_i \sup T_i$ , where  $T_i =$

$$\{t_i \mid \exists c_i, a_i = t_i b_i + (1-t_i) c_i\} . \text{ But } T = \bigcap_i T_i \text{ and } T_i \text{ is}$$

downward closed, this implies  $\sup_i T = \sup(\bigcap_i T_i) = \inf_i(\sup T_i)$ .

Secondly, we show that  $\varphi$  preserves equalizer

$$E \xrightarrow{1} C \begin{array}{c} \xrightarrow{f} \\ \xrightarrow{g} \end{array} B$$

$C, B$  are convex sets,  $E$  is the equalizer of  $f, g$  in  $\mathcal{C}$ . We want to see that the convex metric on  $E$  is the induced convex metric from  $C$ , i.e. if  $e_1, e_2 \in E$ , then  $\text{dist}_C(e_1, e_2) = \text{dist}_E(e_1, e_2)$ .

Now if  $e_1 = te_2 + (1-t)c$  for some  $c \in C$ , then for any  $s < t$ , let  $w = \frac{t-s}{1-s}e_2 + \frac{1-t}{1-s}c$ , since  $tf(e_2) + (1-t)f(c) = f(e_1) = g(e_1) = tg(e_2) + (1-t)g(c)$ , in Lemma 2, let  $a = f(e_2) = g(e_2)$ ,  $b_1 = f(c)$  and  $b_2 = g(c)$ , then  $f(w) = \frac{t-s}{1-s}f(e_2) + \frac{1-t}{1-s}f(c) = \frac{t-s}{1-s}g(e_2) + \frac{1-t}{1-s}g(c) = g(w)$ . Thus  $w \in E$ , and  $\text{dist}_C(e_1, e_2) = \text{dist}_E(e_1, e_2)$ .

Corollary.  $\varphi : \mathcal{C} \longrightarrow \mathcal{V}\text{-cat}$  has a left adjoint.

Proof.  $X$  is a generalized metric space and  $C$  is a convex set,  $f$  is any Lip 1 function from  $X$  to  $\varphi C$ . Because the cardinal number of the convex sets generated by  $f(x)$ ,  $x \in X$ , is bounded, we could take one copy of each isomorphic class of such convex subsets to get a solution set of  $X$ . The above theorem shows that  $\varphi$  is continuous, therefore  $\varphi$  has a left adjoint.

§ 5. The relation between posets and  $\mathcal{U}$ -categories.

Let  $\mathcal{2}$  be the two elements poset.  $\mathcal{2}$  is a cartesian closed category.

Definition 1. Define  $i: \mathcal{2} \longrightarrow \mathcal{U} = ([0, \infty], \geq)$  by  $i(0) = \infty$ ,  $i(1) = 0$ .

Proposition 1.  $i$  is a strict closed functor (strict in the sense that  $i$  preserves tensor product, i.e. for any  $x, y \in \mathcal{2}$ ,  $i(x) + i(y) = i(x \wedge y)$ .)

Proof.  $i$  is order preserving, so it is a functor. We have to check that for any  $x, y \in \mathcal{2}$ ,  $i(x) + i(y) = i(x \wedge y)$ .

If  $x$  or  $y$  is 0, then  $i(x) + i(y) = \infty$ , and  $i(x \wedge y) = i(0) = \infty$ , the equality holds;

if both  $x = 1$  and  $y = 1$ , then  $i(x \wedge y) = i(1) = 0$ , but  $i(x) + i(y) = 0 + 0 = 0$ , the equality still holds.

Remark. The category of posets and order preserving maps is the category of categories enriched in  $\mathcal{2}$ . This closed functor  $i$  induces a functor from Posets to  $\mathcal{U}$ -cat. If  $P$  is a poset, then we have a generalized metric space  $P$  with metric

$$P(p_1, p_2) = \begin{cases} 0, & \text{if } p_1 \leq p_2; \\ \infty, & \text{otherwise.} \end{cases}$$

Definition 2. Define  $\pi_0: \mathcal{U} \longrightarrow \mathcal{2}$  by  $\pi_0(r) = \begin{cases} 1, & r < \infty; \\ 0, & r = \infty. \end{cases}$

Define  $\Gamma: \mathcal{U} \longrightarrow \mathcal{2}$  by  $\Gamma(r) = \begin{cases} 1, & r = 0; \\ 0, & r > 0. \end{cases}$

Proposition 2. Both  $\pi_0$  and  $\Gamma$  are strict closed functors.

$\pi_0$  is left adjoint to  $i$ , and  $\Gamma$  is right adjoint to  $i$ .

Proof. Easy to see  $\pi_0$  and  $\Gamma$  are order preserving.

For  $r_1, r_2 \in \mathcal{U}$ ,

if  $r_1$  or  $r_2$  is equal to  $\infty$ , then  $r_1+r_2 = \infty$ ,  $\pi_0(r_1+r_2) = 0$

and  $\pi_0(r_1) \wedge \pi_0(r_2) = 0$ , the equality

$$\pi_0(r_1+r_2) = \pi_0(r_1) \wedge \pi_0(r_2) \quad \text{holds;}$$

if both  $r_1$  and  $r_2$  are finite, then  $r_1+r_2$  is finite,  $\pi_0(r_1+r_2)=1$

and  $\pi_0(r_1) = \pi_0(r_2) = 1$ , the above equality still holds. Thus,

$\pi_0$  is a strict closed functor.

If  $r_1$  or  $r_2$  is not 0, then  $\Gamma(r_1+r_2) = 0$ , and  $\Gamma(r_1) \wedge \Gamma(r_2) = 0$ ,

the equality  $\Gamma(r_1+r_2) = \Gamma(r_1) \wedge \Gamma(r_2)$  holds;

if both  $r_1$  and  $r_2$  are 0, then  $\Gamma(r_1+r_2) = \Gamma(0) = 1$ , and

$\Gamma(r_1) \wedge \Gamma(r_2) = 1 \wedge 1 = 1$ , the equality still holds. Thus,  $\Gamma$  is

a strict closed functor.

It is not hard to check that for any  $r \in \mathcal{U}$ ,  $x \in \mathcal{Q}$ , the following two bijective equivalences hold

$$\frac{\pi_0(r) \leq x}{r \geq i(x)} \quad \text{and} \quad \frac{i(x) \geq r}{x \leq \Gamma(r)}$$

this shows that  $\pi_0$  is left adjoint to  $i$  and  $\Gamma$  is right adjoint to  $i$ .

Proposition 3.  $\Gamma$  induces the usual underlying poset structure on metric spaces and  $\pi_0$  induces a "face structure" (so called because of the observation that the usual face structure of

a convex set in fact depends only on the associated metric).

Proof.  $\Gamma$  and  $\Pi_0$  induce functors from  $\mathcal{U}$ -cat to Posets, i.e. they induce two different order relations on a generalized metric space. Let  $X$  be a generalized metric space,  $x_1, x_2 \in X$ .  $X$  as a poset with the order induced by  $\Gamma$  is  $x_1 \leq x_2$  iff  $\Gamma(X(x_1, x_2)) = 1$ , i.e.  $X(x_1, x_2) = 0$ . Thus, this is the usual underlying poset structure on metric space  $X$ .

Similarly,  $X$  as a poset with the order induced by  $\Pi_0$  is  $x_1 \leq x_2$  iff  $\Pi_0(X(x_1, x_2)) = 1$ , i.e.  $X(x_1, x_2) < \infty$ . This order relation on finitely generated free convex sets (simplices) will give the usual face structure. We will discuss this at the end of chapter 2, after some work on the topology induced by this generalized metric on a convex set.

CHAPTER 2. CATEGORY  $\mathcal{P}(\mathcal{M})$ .

In this chapter, we are going to discuss that the category  $\mathcal{P}(\mathcal{M})$  of measurable spaces and random maps is the Kleisly category of a monad  $\mathcal{P}$  on the category  $\mathcal{M}$  of measurable spaces and measurable functions, and the properties of  $\mathcal{P}(\mathcal{M})$ . The properties of the topology induced by the convex metric on  $\mathcal{P}(\mathcal{M})(X, Y)$ , for each pair of measurable spaces  $(X, \mathcal{A})$ ,  $(Y, \mathcal{B})$ , also will be discussed here.

§1. Category  $\mathcal{P}(\mathcal{M})$  and its properties.

Let  $\mathcal{M}$  be the category of measurable spaces and measurable functions.

Definition 1.  $(X, \mathcal{A})$  is a measurable space, where  $X$  is a set and  $\mathcal{A}$  is a  $\sigma$ -algebra of  $X$ . For each  $A \in \mathcal{A}$ , we have an evaluation function  $\varepsilon_X(-, A)$  from  $\mathcal{P}(X)$ , the set of probability measures on  $X$ , to  $[0, 1]$ , if  $p \in \mathcal{P}(X)$ , then  $\varepsilon_X(p, A) = p(A)$ .

Definition 2. Let  $(X, \mathcal{A})$  be a measurable space, define  $\mathcal{P}(X) =$  the probability measures on  $X$  with the smallest  $\sigma$ -algebra such that  $\varepsilon_X(-, A)$  is measurable for all  $A \in \mathcal{A}$ .

If  $(Y, \mathcal{B})$  is another measurable space and  $f: X \rightarrow Y$  is a measurable function, define  $\mathcal{P}(f): \mathcal{P}(X) \rightarrow \mathcal{P}(Y)$  by  $\mathcal{P}(f)(p) = p \circ f^{-1}$ , where  $p \in \mathcal{P}(X)$ , i.e. if  $B \in \mathcal{B}$ , then  $p \circ f^{-1}(B) = p(f^{-1}(B))$ .

It is easy to check that  $\mathcal{P}$  is a functor from  $\mathcal{M}$  to  $\mathcal{M}$ .

Lemma.  $(X, \mathcal{A})$ ,  $(Y, \mathcal{B})$  are two measurable spaces. If  $f: X \rightarrow Y$  is a measurable function,  $g$  is a measurable function from  $Y$  to  $\mathbb{R}$ , where  $\mathbb{R}$  is the reals, and  $p$  is any probability measure on  $X$ , then

$$\int_Y g(y) d(p \circ f^{-1}) = \int_X g \circ f(x) dp = \int_X g(f(x)) dp.$$

Proof. See E.Hewitt, Real and abstract analysis, p.180, Theorem (12.46).

Proposition 1.  $\mathcal{P}$  is a monad

Proof. We have to define two natural transformations,  $\eta: 1_{\mathcal{M}} \rightarrow \mathcal{P}$  and  $\mu: \mathcal{P}\mathcal{P} \rightarrow \mathcal{P}$ , and have to show that they are satisfying some commutative diagrams.

Let  $(X, \mathcal{A})$ ,  $(Y, \mathcal{B})$  be two measurable spaces.

Define  $\eta_X: X \rightarrow \mathcal{P}(X)$  by  $\eta_X(x) = \delta_x$ , for each  $x \in X$ ,  $\delta_x$  is the probability measure concentrated on  $x$ , i.e. for  $A \in \mathcal{A}$ ,

$$\delta_x(A) = \begin{cases} 1, & x \in A; \\ 0, & x \notin A. \end{cases}$$

If  $f: X \rightarrow Y$  is a measurable function, the following diagram is commutative,

$$\begin{array}{ccc} X & \xrightarrow{\eta_X} & \mathcal{P}(X) \\ \downarrow f & & \downarrow \mathcal{P}(f) \\ Y & \xrightarrow{\eta_Y} & \mathcal{P}(Y) \end{array}$$

since for  $x \in X$ ,  $B \in \mathcal{B}$ ,  $\mathcal{P}(f) \circ \eta_X(x)(B) = \mathcal{P}(f)(\delta_x(B)) = \delta_x(f^{-1}(B))$

$$= \begin{cases} 1, & x \in f^{-1}(B); \\ 0, & x \notin f^{-1}(B). \end{cases} = \begin{cases} 1, & f(x) \in B; \\ 0, & f(x) \notin B. \end{cases} = \delta_{f(x)}(B),$$

but  $\eta_Y \circ f(x)(B) = \delta_{f(x)}(B)$ . Thus,  $\eta$  is natural.

Define  $\mu_X: \mathcal{P}\mathcal{P}(X) \rightarrow \mathcal{P}(X)$ , for  $q \in \mathcal{P}\mathcal{P}(X)$ ,  $A \in \mathcal{A}$ ,

$$\mu_X(q)(A) = \int_{\mathcal{P}(X)} \mathcal{E}_X(p, A) dq, \text{ note that, } q \text{ is a probability measure on } \mathcal{P}(X).$$

measure on  $\mathcal{P}(X)$ .

Let  $q \in \mathcal{P}\mathcal{P}(X)$ ,  $B \in \mathcal{B}$ , then in the following diagram,

$$\begin{array}{ccc} \mathcal{P}\mathcal{P}(X) & \xrightarrow{\mu_X} & \mathcal{P}(X) \\ \downarrow \mathcal{P}\mathcal{P}(f) & & \downarrow \mathcal{P}(f) \\ \mathcal{P}\mathcal{P}(Y) & \xrightarrow{\mu_Y} & \mathcal{P}(Y) \end{array}$$

the left side:  $\mu_Y \circ \mathcal{P}\mathcal{P}(f)(q)(B)$

$$\begin{aligned} &= \mu_Y(q \circ \mathcal{P}(f)^{-1})(B) \\ &= \int_{\bar{p} \in \mathcal{P}(Y)} \varepsilon_Y(\bar{p}, B) d(q \circ \mathcal{P}(f)^{-1}) \\ &= \int_{p \in \mathcal{P}(X)} \varepsilon_Y(\mathcal{P}(f)(p), B) dq \quad (\text{by Lemma}) \\ &= \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, f^{-1}(B)) dq, \quad (\text{since } \varepsilon_Y(\mathcal{P}(f)(p), B) \\ & \qquad \qquad \qquad = \varepsilon_X(p, f^{-1}(B))) \end{aligned}$$

the right side:  $\mathcal{P}(f) \circ \mu_X(q)(B)$

$$\begin{aligned} &= \mu_X(q)(f^{-1}(B)) \\ &= \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, f^{-1}(B)) dq. \end{aligned}$$

thus, left = right, i.e.  $\mu$  is natural.

Diagram 1.

$$\begin{array}{ccccc} & & \mathcal{P}\mathcal{P}(X) & & \\ & \nearrow \eta_{\mathcal{P}(X)} & & \nwarrow \mathcal{P}(\eta_X) & \\ \mathcal{P}(X) & & & & \mathcal{P}(X) \\ & \searrow 1_{\mathcal{P}(X)} & \downarrow \mu_X & \swarrow 1_{\mathcal{P}(X)} & \\ & & \mathcal{P}(X) & & \end{array}$$

let  $p \in \mathcal{P}(X)$ ,  $A \in \mathcal{A}$ ,

the left triangle:

$$\begin{aligned} \mu_X \circ \mathcal{I}_{\mathcal{P}(X)}(p)(A) &= \mu_X \circ \delta_p(A) \\ &= \int_{q \in \mathcal{P}(X)} \varepsilon_{X(q,A)} d\delta_p \\ &= p(A) \end{aligned}$$

thus,  $\mu_X \circ \mathcal{I}_{\mathcal{P}(X)}(p) = p$ , i.e.  $\mu_X \circ \mathcal{I}_{\mathcal{P}(X)} = 1_{\mathcal{P}(X)}$ ;

the right triangle:

$$\begin{aligned} \mu_X \circ \mathcal{P}(\mathcal{I}_X)(p)(A) &= \mu_X(\mathcal{P} \circ \mathcal{I}_X^{-1})(A) \\ &= \int_{q \in \mathcal{P}(X)} \varepsilon_{X(q,A)} d(\mathcal{P} \circ \mathcal{I}_X^{-1}) \\ &= \int_{x \in X} \varepsilon_{X(\mathcal{I}_X(x), A)} dp \quad (\text{by Lemma}) \\ &= p(A), \quad (\text{since } \varepsilon_{X(\mathcal{I}_X(x), A)} = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}) \end{aligned}$$

thus,  $\mu_X \circ \mathcal{P}(\mathcal{I}_X)(p) = p$ , i.e.  $\mu_X \circ \mathcal{P}(\mathcal{I}_X) = 1_{\mathcal{P}(X)}$ .

Diagram 2.

$$\begin{array}{ccc} \mathcal{P}\mathcal{P}\mathcal{P}(X) & \xrightarrow{\mathcal{P}(\mu_X)} & \mathcal{P}\mathcal{P}(X) \\ \downarrow \mu_{\mathcal{P}(X)} & & \downarrow \mu_X \\ \mathcal{P}\mathcal{P}(X) & \xrightarrow{\mu_X} & \mathcal{P}(X) \end{array}$$

let  $q \in \mathcal{P}\mathcal{P}\mathcal{P}(X)$ ,  $A \in \mathcal{A}$ ,

the left side:

$$\begin{aligned}
\mu_X \circ \mu_{\mathcal{P}(X)}(q)(A) &= \mu_X(\mu_{\mathcal{P}(X)}(q))(A) \\
&= \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, A) d\mu_{\mathcal{P}(X)}(q) \\
&= \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, A) \int_{\sigma \in \mathcal{P}\mathcal{P}(X)} \varepsilon_{\mathcal{P}(X)}(\sigma, dp) dq \\
&= \int_{\sigma \in \mathcal{P}\mathcal{P}(X)} \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, A) \varepsilon_{\mathcal{P}(X)}(\sigma, dp) dq ; \\
&= \int_{\sigma \in \mathcal{P}\mathcal{P}(X)} \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, A) d\sigma dq
\end{aligned}$$

the right side:

$$\begin{aligned}
\mu_X \circ \mathcal{P}(\mu_X)(q)(A) &= \mu_X(q \circ \mu_X^{-1})(A) \\
&= \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, A) d(q \circ \mu_X^{-1}) \\
&= \int_{\sigma \in \mathcal{P}\mathcal{P}(X)} \varepsilon_X(\mu_X(\sigma), A) dq \quad (\text{note, } \mu_X: \mathcal{P}\mathcal{P}(X) \longrightarrow \mathcal{P}(X)) \\
&= \int_{\sigma \in \mathcal{P}\mathcal{P}(X)} \mu_X(\sigma)(A) dq \\
&= \int_{\sigma \in \mathcal{P}\mathcal{P}(X)} \int_{p \in \mathcal{P}(X)} \varepsilon_X(p, A) d\sigma dq
\end{aligned}$$

thus, left = right, the square in diagram 2 is commutative.

This completes the proof that  $\mathcal{P}$  is a monad on  $\mathcal{M}$ .

Denote the Kleisli category of monad  $\mathcal{P}$  by  $\mathcal{P}(\mathcal{M})$ .  $\mathcal{P}(\mathcal{M})$  has the same objects as  $\mathcal{M}$  does. If  $(X, \mathcal{A})$  and  $(Y, \mathcal{B})$  are two measurable spaces, a morphism  $f$  from  $X$  to  $Y$  in  $\mathcal{P}(\mathcal{M})$  is a morphism  $f: X \longrightarrow \mathcal{P}(Y)$  in  $\mathcal{M}$ , i.e. for each  $x$  in  $X$ ,  $f(x)$  is a probability measure on  $Y$ ;

for each  $B \in \mathcal{B}$ , the composition  $\varepsilon_Y(-, B) \circ f: X \xrightarrow{f} \mathcal{P}(Y) \xrightarrow{\varepsilon_Y(-, B)} [0, 1]$  is  $\mathcal{A}$ -measurable. Thus,  $f$  is just a probabilistic transformation from  $X$  to  $Y$  ( $f$  is a probabilistic transformation between two measurable spaces  $(X, \mathcal{A})$  and  $(Y, \mathcal{B})$ , if  $f$  is a function from  $X$  to  $\mathcal{P}(Y)$ , such that for each  $x \in X$ ,  $f(x, -)$  is a probability measure on  $(Y, \mathcal{B})$ ; and for each  $B \in \mathcal{B}$ ,  $f(-, B)$  is  $\mathcal{A}$ -measurable). We will call a morphism  $f$  in  $\mathcal{P}(\mathcal{M})$  a random map, denote it as  $X \xrightarrow{f} Y$ ; and for  $x \in X$ ,  $B \in \mathcal{B}$ , denote  $f(x, B)$  for  $f(x)(B)$ .

If  $f: X \rightarrow Y$  and  $g: Y \rightarrow Z$  are two random maps, where  $X, Y$  and  $Z$  have  $\sigma$ -algebras  $\mathcal{A}, \mathcal{B}$ , and  $\mathcal{C}$  respectively, according to the rule of composition in Kleisli category,  $g \circ f = \mu_Z \circ \mathcal{P}(g) \circ f$ :

$$\begin{aligned} X &\xrightarrow{f} \mathcal{P}(Y) \xrightarrow{\mathcal{P}(g)} \mathcal{P}(\mathcal{P}(Z)) \xrightarrow{\mu_Z} \mathcal{P}(Z). \text{ Let } x \in X, C \in \mathcal{C}, \text{ then} \\ g \circ f(x, C) &= \mu_Z \circ \mathcal{P}(g) \circ f(x, C) = \int_{q \in \mathcal{P}(Z)} \varepsilon_Z(q, C) d(f(x, -) \circ g^{-1}) \\ &= \int_{y \in Y} \varepsilon_Z(g(y, -), C) f(x, dy) = \int_Y g(y, C) f(x, dy), \end{aligned}$$

where  $f(x, dy)$  stands for  $df(x, -)$ .

From the adjointness,  $\mathcal{M} \xrightleftharpoons[\mathcal{G}_\mathcal{P}]{\mathcal{F}_\mathcal{P}} \mathcal{P}(\mathcal{M})$  (the notation from MacLane's book, *Categories for the working mathematician*, Springer-Verlag, 1971), any measurable function  $f: (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$  could be regarded as a random map, for  $x \in X$ ,  $B \in \mathcal{B}$ ,

$$f(x, B) = \begin{cases} 1, & f(x) \in B, \\ 0, & f(x) \notin B. \end{cases} \text{ We will call this } f \text{ deterministic .}$$

Proposition 2.  $\mathcal{P}(\mathcal{M})$  is tensored, and the tensor is a bifunctor.

Proof. If  $(X, \mathcal{A}), (Y, \mathcal{B}) \in \mathcal{P}(\mathcal{M})$ , define  $(X, \mathcal{A}) \otimes (Y, \mathcal{B}) = (X \times Y, \mathcal{A} \times \mathcal{B})$ ,

where  $X \times Y$  is the cartesian product of sets  $X$  and  $Y$ ,  $\mathcal{A} \otimes \mathcal{B}$  is the  $\sigma$ -algebra generated by the measurable rectangles  $\{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}$  on  $X \times Y$ .

Let  $f: (X, \mathcal{A}) \rightarrow (X', \mathcal{A}')$ ,  $g: (Y, \mathcal{B}) \rightarrow (Y', \mathcal{B}')$  be two random maps, define  $f \otimes g: (X, \mathcal{A}) \otimes (Y, \mathcal{B}) \rightarrow (X', \mathcal{A}') \otimes (Y', \mathcal{B}')$  by

$f \otimes g((x, y), A' \times B') = f(x, A') \cdot g(y, B')$ , where  $(x, y) \in X \times Y$ ,  $A' \times B'$  is a measurable rectangle in  $\mathcal{A}' \times \mathcal{B}'$ .

1). For each pair of  $(x, y) \in X \times Y$ ,  $f \otimes g((x, y), -)$  is a probability measure on  $(X' \times Y', \mathcal{A}' \times \mathcal{B}')$ :

first,  $f \otimes g((x, y), X' \times Y') = f(x, X') \cdot g(y, Y') = 1$ ;

if  $A' \times B' = \bigcup_{i=1}^{\infty} A'_i \times B'_i$ ,  $A'_i \times B'_i \cap A'_j \times B'_j = \emptyset$  for  $i \neq j$ , then

either  $A'_i \cap A'_j = \emptyset$  or  $B'_i \cap B'_j = \emptyset$ .

$$\begin{aligned} \text{Thus, } f \otimes g((x, y), A' \times B') &= f(x, A') \cdot g(y, B') \\ &= \int_{Y'} \left( \int_{X'} I_{A'} \cdot I_{B'} f(x, dx') \right) g(y, dy') \\ &= \int_{Y'} \left( \int_{X'} I_{A' \times B'} f(x, dx') \right) g(y, dy') \\ &= \int_{Y'} \left( \int_{X'} \sum_{i=1}^{\infty} I_{A'_i \times B'_i} f(x, dx') \right) g(y, dy') \\ &= \sum_{i=1}^{\infty} \int_{Y'} \left( \int_{X'} I_{A'_i} \cdot I_{B'_i} f(x, dx') \right) g(y, dy') \\ &= \sum_{i=1}^{\infty} f(x, A'_i) \cdot g(y, B'_i) \\ &= \sum_{i=1}^{\infty} f \otimes g((x, y), A'_i \times B'_i), \end{aligned}$$

then use Caratheodory extension theorem, extend  $f \otimes g((x, y), -)$  to a probability measure on  $\mathcal{A}' \times \mathcal{B}'$ .

2). For any  $A' \times B'$ ,  $f \otimes g(\_, A' \times B')$  is  $\mathcal{A} \times \mathcal{B}$ -measurable:

we start with that  $f(\_, A')$  is a characteristic function,

$f(\_, A') = I_A$ , for some  $A \in \mathcal{A}$ , then

$$\begin{aligned} f \otimes g((x, y), A' \times B') &= f(x, A') \cdot g(y, B') \\ &= I_A(x) \cdot g(y, B') \end{aligned}$$

if  $u$  is a measurable set in  $[0, 1]$ , and let  $A^c = X \setminus A$ ,

$$(I_A(x) \cdot g(y, B'))^{-1}(u) = \begin{cases} A \times g^{-1}(y, B')(u), & \text{if } 0 \notin u; \\ A^c \times Y \cup A \times g^{-1}(y, B')(u), & \text{if } 0 \in u, \end{cases}$$

thus,  $I_A(x) \cdot g(y, B')$  is  $\mathcal{A} \times \mathcal{B}$ -measurable;

if  $f(\_, A') = \sum_{i=1}^n a_i I_{A_i}$  is a step function,  $a_i \in \mathbb{R}$  the reals,

$$\begin{aligned} \text{then } f \otimes g((x, y), A' \times B') &= \left( \sum_{i=1}^n a_i I_{A_i}(x) \right) \cdot g(y, B') \\ &= \sum_{i=1}^n a_i (I_{A_i}(x) \cdot g(y, B')) \end{aligned}$$

since a finite sum of measurable functions is measurable, thus

in this case,  $f \otimes g(\_, A' \times B')$  is  $\mathcal{A} \times \mathcal{B}$ -measurable;

now  $f(\_, A') = \lim_{n \rightarrow \infty} s_n$ , where  $\{s_n\}$  is a sequence of increasing step

functions, then  $f \otimes g((x, y), A' \times B') = f(x, A') \cdot g(y, B')$

$$= \left( \lim_{n \rightarrow \infty} s_n \right) \cdot g(y, B')$$

$$= \lim_{n \rightarrow \infty} s_n \cdot g(y, B')$$

since for each  $n$ ,  $s_n \cdot g(y, B')$  is a  $\mathcal{A} \times \mathcal{B}$ -measurable function, and

the limit of a sequence of measurable functions is measurable,

thus  $f \otimes g(\_, A' \times B')$  is  $\mathcal{A} \times \mathcal{B}$ -measurable in this case.

Now from the proof of Caratheodory extension theorem, we get that

for any  $C \in \mathcal{A} \times \mathcal{B}$ ,  $f \otimes g(\_, C)$  is  $\mathcal{A} \times \mathcal{B}$ -measurable.

To see that tensor  $\otimes : \mathcal{P}(M) \times \mathcal{P}(M) \longrightarrow \mathcal{P}(M)$  is a bifunctor, let

$$(X, \mathcal{A}) \xrightarrow{f} (X', \mathcal{A}') \xrightarrow{f'} (X'', \mathcal{A}'' )$$

$$(Y, \mathcal{B}) \xrightarrow{g} (Y', \mathcal{B}') \xrightarrow{g'} (Y'', \mathcal{B}'' )$$

be random maps, we have to show that the following diagram is commutative, i.e. to show that  $(f' \circ f) \otimes (g' \circ g) = f' \otimes g' \circ f \otimes g$

$$\begin{array}{ccc} X \times Y & \xrightarrow{(f' \circ f) \otimes (g' \circ g)} & X'' \times Y'' \\ & \searrow f \otimes g & \nearrow f' \otimes g' \\ & X' \times Y' & \end{array}$$

let  $(x, y) \in X \times Y$ ,  $A'' \times B''$  be a measurable rectangle in  $\mathcal{A}'' \times \mathcal{B}''$ ,

$$\begin{aligned} & f' \otimes g' \circ f \otimes g((x, y), A'' \times B'') \\ &= \int_{X' \times Y'} f' \otimes g'((x', y'), A'' \times B'') f \otimes g((x, y), d(x', y')) \\ &= \int_{X' \times Y'} f'(x', A'') g'(y', B'') f(x, dx') g(y, dy') \\ &= \int_{X'} f'(x', A'') f(x, dx') \int_{Y'} g'(y', B'') g(y, dy') \quad (\text{by Fubini theorem}) \\ &= f' \circ f(x, A'') g' \circ g(y, B'') \end{aligned}$$

therefore,  $(f' \circ f) \otimes (g' \circ g) = f' \otimes g' \circ f \otimes g$ .

We use  $\mathcal{P}(M)(X, Y)$  to denote the set of random maps from  $(X, \mathcal{A})$  to  $(Y, \mathcal{B})$ . For  $f, g \in \mathcal{P}(M)(X, Y)$ ,  $0 \leq t \leq 1$ ,  $x \in X$ , and  $B \in \mathcal{B}$ , define  $(tf + (1-t)g)(x, B) = tf(x, B) + (1-t)g(x, B)$ . It is easy to see that  $tf + (1-t)g \in \mathcal{P}(M)(X, Y)$ , thus  $\mathcal{P}(M)(X, Y)$  is a convex set.

Proposition 3. The tensor is biconvex, in the sense that if  $f, f_1, f_2 : (X, \mathcal{A}) \longrightarrow (Y, \mathcal{B})$ ,  $g, g_1, g_2 : (X', \mathcal{A}') \longrightarrow (Y', \mathcal{B}')$  are random maps, then for any  $0 \leq t \leq 1$ ,

$$f \otimes (tg_1 + (1-t)g_2) = tf \otimes g_1 + (1-t)f \otimes g_2$$

and

$$(tf_1 + (1-t)f_2) \otimes g = tf_1 \otimes g + (1-t)f_2 \otimes g.$$

Proof. Let  $(x, x') \in X \times X'$ ,  $B \times B' \in \mathcal{B} \times \mathcal{B}'$ ,

$$\begin{aligned} & f \otimes (tg_1 + (1-t)g_2)((x, x'), B \times B') \\ &= f(x, B)(tg_1 + (1-t)g_2)(x', B') \\ &= tf(x, B)g_1(x', B') + (1-t)f(x, B)g_2(x', B') \\ &= tf \otimes g_1((x, x'), B \times B') + (1-t)f \otimes g_2((x, x'), B \times B') \end{aligned}$$

similarly

$$\begin{aligned} & (tf_1 + (1-t)f_2) \otimes g((x, x'), B \times B') \\ &= (tf_1 \otimes g + (1-t)f_2 \otimes g)((x, x'), B \times B') \end{aligned}$$

then use Caratheodory extension theorem to extend the result to any measurable set  $C \in \mathcal{B} \times \mathcal{B}'$ . Thus, we have that  $\otimes$  is biconvex.

Proposition 4. The finite product  $(\prod_{i=1}^n X_i, \prod_{i=1}^n \mathcal{A}_i)$  of  $(X_i, \mathcal{A}_i)$  with projections  $\pi_i: \prod_{i=1}^n X_i \rightarrow X_i$  in  $\mathcal{M}$  is a weak product in  $\mathcal{P}(\mathcal{M})$ .

The weak product is in the sense that for any  $(Y, \mathcal{B})$  and  $f_i: Y \rightarrow X_i$ , there is a  $f: Y \rightarrow \prod_{i=1}^n X_i$ , such that  $\pi_i \circ f = f_i$  ( $f$  is not necessary uniquely determined by the latter equations).

Proof. Define  $f = \bigotimes_{i=1}^n f_i \circ \Delta_Y$ , where  $\Delta_Y: Y \rightarrow Y \times Y \times \dots \times Y$  is the diagonal map viewed as a random map, i.e. for  $y \in Y$ ,  $B_i \in \mathcal{B}$ ,  $i=1, \dots, n$ ,

$$\Delta_Y(y, B_1 \times \dots \times B_n) = \begin{cases} 1, & \text{if } y \in B_1 \cap \dots \cap B_n, \\ 0, & \text{otherwise.} \end{cases}$$

Since the unit of tensor is the terminal object in  $\mathcal{P}(\mathcal{M})$ , we have the projection  $\pi_i = (\prod_{i=1}^n X_i \rightarrow 1 \times \dots \times X_i \times \dots \times 1)$ , this gives that  $\pi_i \circ f = f_i$ .

Note that,  $\Delta_Y$  is not natural in  $\mathcal{P}(\mathcal{M})$ , unless  $f: X \rightarrow Y$  is a measurable function (i.e.  $f$  is deterministic). In the following diagram

$$\begin{array}{ccc} X & \xrightarrow{\Delta_X} & X \times X \\ \downarrow f & & \downarrow f \otimes f \\ Y & \xrightarrow{\Delta_Y} & Y \times Y \end{array}$$

let  $x \in X$  and  $B_1, B_2 \in \mathcal{B}$  the measurable sets in  $Y$ , then

$$\begin{aligned} \Delta_Y \circ f(x, B_1 \times B_2) &= \int_Y \Delta_Y(y, B_1 \times B_2) f(x, dy) = f(x, B_1 \cap B_2) \\ f \otimes f \circ \Delta_X(x, B_1 \times B_2) &= \int_{X \times X} f \otimes f((x_1, x_2), B_1 \times B_2) \Delta_X(x, d(x_1, x_2)) \\ &= f(x, B_1) f(x, B_2) \end{aligned}$$

but  $f(x, B_1 \cap B_2)$  may not be equal to  $f(x, B_1) f(x, B_2)$ , such as  $X = \{x\}$ ,  $Y = \{a, b\}$ ,  $f(x, a) = \frac{1}{2} = f(x, b)$ , then  $f(x, \{a\} \cap \{b\}) = f(x, \emptyset) = 0$ , but  $f(x, \{a\}) f(x, \{b\}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ .

Proposition 5.  $\mathcal{P}(\mathcal{M})$  is enriched in  $\mathbb{C}$ .

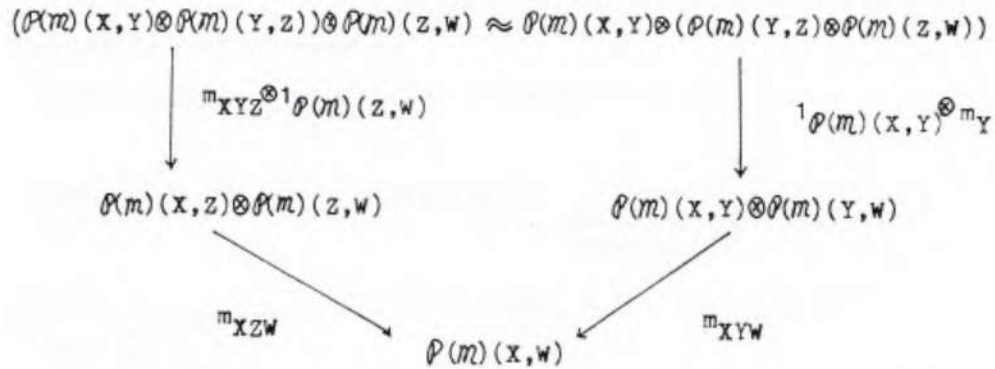
For each pair of measurable spaces  $(X, \mathcal{A})$ ,  $(Y, \mathcal{B})$ ,  $\mathcal{P}(\mathcal{M})(X, Y)$  is a convex set.

Define  $j_X: 1 \rightarrow \mathcal{P}(\mathcal{M})(X, X)$ ,  $1 \mapsto 1_X$ ;

$$\begin{aligned} m_{XYZ}: \mathcal{P}(\mathcal{M})(X, Y) \otimes \mathcal{P}(\mathcal{M})(Y, Z) &\longrightarrow \mathcal{P}(\mathcal{M})(X, Z) \\ \sum_{i=1}^n r_i f_i \otimes g_i &\longmapsto \sum_{i=1}^n r_i g_i \circ f_i \end{aligned}$$

where  $f_i \in \mathcal{P}(\mathcal{M})(X, Y)$ ,  $g_i \in \mathcal{P}(\mathcal{M})(Y, Z)$ ,  $r_i \in \mathbb{K}$  and  $\sum_{i=1}^n r_i = 1$ ,  $\otimes$  is the tensor in  $\mathbb{C}$ . Because  $\sum_{i=1}^n r_i = 1$  and  $g_i \circ f_i \in \mathcal{P}(\mathcal{M})(X, Z)$ , so  $\sum_{i=1}^n r_i g_i \circ f_i \in \mathcal{P}(\mathcal{M})(X, Z)$ . and both  $j_X$ ,  $m_{XYZ}$  are affine linear.

The associative law:



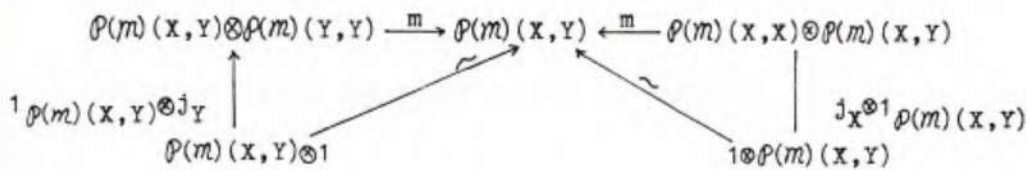
let  $r_i \in K$ ,  $\sum_{i=1}^n r_i = 1$ ,  $f_i \in \mathcal{P}(m)(X,Y)$ ,  $g_i \in \mathcal{P}(m)(Y,Z)$  and  $h_i \in \mathcal{P}(m)(Z,W)$ .

then

$$\begin{array}{ccc}
 \sum_{i=1}^n r_i (f_i \otimes g_i) \otimes h_i & = & \sum_{i=1}^n r_i f_i \otimes (g_i \otimes h_i) \\
 \downarrow & & \downarrow \\
 \sum_{i=1}^n r_i (g_i \circ f_i) \otimes h_i & & \sum_{i=1}^n r_i f_i \otimes (h_i \circ g_i) \\
 \swarrow & & \swarrow \\
 & \sum_{i=1}^n r_i h_i \circ g_i \circ f_i & 
 \end{array}$$

thus, the diagram is commutative.

The identity law:



let  $f \in \mathcal{P}(m)(X,Y)$ , since  $f \circ 1_X = f = 1_Y \circ f$ , thus both triangles are commutative.

Note that, actually  $\mathcal{P}(m)$  is enriched in  $\mathcal{U}$ -cat too, since there is a closed functor from  $\mathcal{C}$  to  $\mathcal{U}$ -cat.

One of the ideas to understand the distance between two random maps  $f, g: (X, \mathcal{A}) \rightarrow (Y, \mathcal{B})$ , which will be used later in discussing completeness is

Proposition 6.  $\sup_{\substack{x \in X \\ B \in \mathcal{B}}} |f(x, B) - g(x, B)| < \text{dist}(f, g).$

Proof. Let us denote  $\text{dist}(f, g) = \begin{matrix} -\log \sup t \\ \exists t \in [0, 1], \exists h \in \mathcal{P}(m)(X, Y) \\ f = tg + (1-t)h \end{matrix} = d.$

If  $\sup_{\substack{t \in [0, 1], \exists h \in \mathcal{P}(m)(X, Y) \\ f = tg + (1-t)h}} t = t_0$ , then  $-\log t_0 = d$ , or  $t_0 = e^{-d}$ .

For any  $\delta$ ,  $0 < \delta < t_0$ ,  $\exists t > t_0 - \delta$ ,  $\exists h \in \mathcal{P}(m)(X, Y)$ , such that  $f = tg + (1-t)h$ , i.e.  $f(x, B) = tg(x, B) + (1-t)h(x, B)$ , for  $x \in X$ ,  $B \in \mathcal{B}$ ,

$$\begin{aligned} \text{So, } |f(x, B) - g(x, B)| &= (1-t)|h(x, B) - g(x, B)| \\ &\leq 1-t \quad (\text{since } 0 \leq h(x, B), g(x, B) \leq 1) \\ &\leq 1 - t_0 + \delta \end{aligned}$$

this implies that  $|f(x, B) - g(x, B)| \leq 1 - t_0$  for any  $x \in X$ ,  $B \in \mathcal{B}$ ,

thus,  $\sup_{\substack{x \in X \\ B \in \mathcal{B}}} |f(x, B) - g(x, B)| \leq 1 - t_0$ .

Define  $y(x) = 1 - e^{-x} - x$ , the derivative of  $y(x)$ ,  $y'(x) = e^{-x} - 1 < 0$ , as  $x > 0$ ; because also  $y(0) = 0$ , so  $1 - e^{-x} < x$ , specially for  $x = d$ ,  $1 - e^{-d} < d$ , i.e.  $1 - t_0 = 1 - e^{-d} < d$ .

This completes the proof of the proposition.

§ 2. The covariant and contravariant topologies of  $\mathcal{P}(M)(X, Y)$ .

Definition 1.  $C$  is a convex set, the covariant open ball in  $C$  with center  $a$  and radius  $r$  is  $\{c \in C \mid \text{dist}(a, c) < r\} = B(a, r)$ , where  $a \in C$ ,  $r \in K$ .

The contravariant open ball in  $C$  with center  $a$  and radius  $r$  is  $\{c \in C \mid \text{dist}(c, a) < r\} = B(r, a)$ .

Proposition 1.  $\{B(r, a) \mid a \in C, r \in K\}$  form a basis.

Proof. Let  $c \in B(r, a) \cap B(s, b)$ ,  $\text{dist}(c, a) = r_1 < r$  and  $\text{dist}(c, b) = s_1 < s$ . Let  $\delta = \min\{r - r_1, s - s_1\}$ . If  $x \in B(\delta, c)$ , that is  $\text{dist}(x, c) < \delta$ , then  $\text{dist}(x, a) \leq \text{dist}(x, c) + \text{dist}(c, a) < \delta + r_1 \leq r - r_1 + r_1 = r$ , and  $\text{dist}(x, b) \leq \text{dist}(x, c) + \text{dist}(c, b) < \delta + s_1 \leq s - s_1 + s_1 = s$ . Thus,  $B(\delta, c) \subseteq B(r, a) \cap B(s, b)$ .

Using a similar argument, we can show that the covariant open balls also form a basis for covariant topology induced by the metric.

Remark. Let  $f, g, h \in \mathcal{P}(M)(X, Y)$ . The convex combination  $f = tg + (1-t)h$  implies that  $f(x, B) = tg(x, B) + (1-t)h(x, B)$ , for all  $x \in X$  and  $B \in \mathcal{B}$ . If  $f$  is deterministic, i.e.  $f(x, B) = \begin{cases} 1, & f(x) \in B \\ 0, & f(x) \notin B \end{cases}$ , then this forces  $g(x, B) = h(x, B) = \begin{cases} 1, & f(x) \in B \\ 0, & f(x) \notin B \end{cases}$ , that is  $g = h = f$ . Thus, the covariant open ball  $B(f, r) = \{f\}$ , or  $f$  is isolated in the covariant topology. However, the deterministic map is not isolated in the contravariant topology. This is the reason that we feel more interested in the contravariant one.

Examples.

First note, if  $b = ta + (1-t)c$ ,  $0 < t < 1$ ,  $a, b, c$  are reals,  $a \neq b$ , then either  $a < b < c$  or  $a > b > c$ .

Example 1. The unit interval  $[0, 1]$ .

If  $b = ta + (1-t)c$ ,

case 1.  $0 \leq a < b < c \leq 1$ ,  $\frac{a \quad b \quad c}{0 \quad \quad \quad 1}$ , then  $t = \frac{c-b}{c-a}$ ;

case 2.  $0 \leq c < b < a \leq 1$ ,  $\frac{c \quad b \quad a}{0 \quad \quad \quad 1}$ , then  $t = \frac{b-c}{a-c} = \frac{c-b}{c-a}$ .

Since the derivative of function  $y(c) = \frac{c-b}{c-a}$  is  $y'(c) = \frac{b-a}{(c-a)^2}$ ,

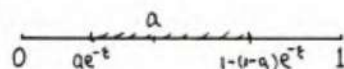
so  $y(c)$  is increasing as  $b > a$ , i.e. in case 1;  $y(c)$  is decreasing as  $b < a$ , i.e. in case 2.

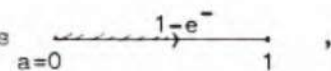
$$\begin{aligned} \text{Thus, } \text{dist}(b, a) &= \begin{matrix} -\log \sup_{\substack{t \in [0, 1], \\ b = ta + (1-t)c}} t \\ \exists t \in [0, 1], \exists c \in [0, 1] \\ b = ta + (1-t)c \end{matrix} = -\log \sup_{c \in [0, 1]} \frac{c-b}{c-a} \\ &= \begin{cases} -\log \frac{1-b}{1-a}, & \text{if } b > a, \\ -\log \frac{b}{a}, & \text{if } b < a. \end{cases} \end{aligned}$$

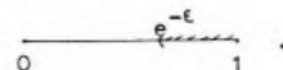
The open ball in  $[0, 1]$  with center  $a$  and radius  $\xi$  (contravariant)

$$\begin{aligned} \text{is } B(\xi, a) &= \{ b \in [0, 1] \mid \text{dist}(b, a) < \xi \} \\ &= \left\{ b \in [0, 1] \mid \begin{array}{l} -\log \frac{1-b}{1-a} < \xi, \quad b > a \\ -\log \frac{b}{a} < \xi, \quad b \leq a \end{array} \right\} \\ &= \left\{ b \in [0, 1] \mid \begin{array}{ll} b < 1 - (1-a)e^{-\xi}, & b > a \\ b > ae^{-\xi}, & b \leq a \end{array} \right\} \end{aligned}$$

$B(\xi, a)$  is as in the picture:

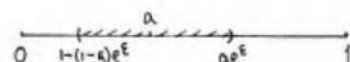


if  $a = 0$ ,  $B(\varepsilon, 0) = \{b \in [0, 1] \mid b < 1 - e^{-\varepsilon}\}$  is as  ,

if  $a = 1$ ,  $B(\varepsilon, 1) = \{b \in [0, 1] \mid b > e^{-\varepsilon}\}$  is as  .

The open ball in  $[0, 1]$  with center  $a$  and radius  $\varepsilon$  (covariant) is

$$\begin{aligned} B(a, \varepsilon) &= \{b \in [0, 1] \mid \text{dist}(a, b) < \varepsilon\} \\ &= \left\{ b \in [0, 1] \mid \begin{array}{l} -\log \frac{1-a}{1-b} < \varepsilon, \quad b < a \\ -\log \frac{a}{b} < \varepsilon, \quad b \geq a \end{array} \right\} \\ &= \left\{ b \in [0, 1] \mid \begin{array}{l} b > 1 - (1-a)e^\varepsilon, \quad b < a \\ b < ae^\varepsilon, \quad b \geq a \end{array} \right\} \end{aligned}$$

as in the picture 

if  $a = 0$ ,  $ae^\varepsilon = 0$ , so  $B(0, \varepsilon) = \{b \in [0, 1] \mid b \leq 0\} = \{0\}$  ,

if  $a = 1$ ,  $1 - (1-a)e^\varepsilon = 1$ , so  $B(1, \varepsilon) = \{b \in [0, 1] \mid b \geq 1\} = \{1\}$  .

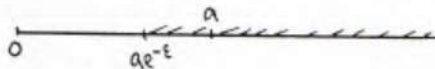
Remark. a). This works for any closed interval  $[a, b]$ ,  $-\infty < a < b < \infty$   
 b). The topologies (contravariant, covariant) induced by this metric  $\text{dist}$  on  $(0, 1)$  are equivalent to the usual topology on  $(0, 1)$ .

Example 2.  $[0, \infty)$ .

From Example 1,  $y(c) = \frac{c-b}{c-a}$  is increasing as  $b > a$ , decreasing as  $b < a$ . Thus, as  $b > a$ ,  $\sup_{c \in [0, \infty)} \frac{c-b}{c-a} = \lim_{c \rightarrow \infty} \frac{c-b}{c-a} = \lim_{c \rightarrow \infty} (1 - \frac{b-a}{c-a}) = 1$ ; and as  $b < a$ ,  $\sup_{c \in [0, \infty)} \frac{c-b}{c-a} = \frac{b}{a}$  .

$$\begin{aligned}
\text{Thus, } B(\varepsilon, a) &= \{b \in [0, \infty) \mid \text{dist}(b, a) < \varepsilon\} \\
&= \left\{ b \in [0, \infty) \mid -\log \sup_{c \in [0, \infty)} \frac{c-b}{c-a} < \varepsilon \right\} \\
&= \left\{ b \in [0, \infty) \mid \begin{array}{l} -\log 1 = 0 < \varepsilon, \quad b > a \\ -\log \frac{b}{a} < \varepsilon, \quad b < a \end{array} \right\} \\
&= \{b \in [0, \infty) \mid b > ae^{-\varepsilon}\}
\end{aligned}$$

$B(\varepsilon, a)$  is as in the picture:

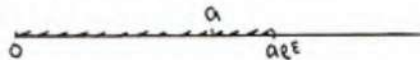


if  $a = 0$ , then  $ae^{-\varepsilon} = 0$ , thus  $B(\varepsilon, 0) = \{b \in [0, \infty) \mid b > 0\} = [0, \infty)$ .

The covariant open ball

$$\begin{aligned}
B(a, \varepsilon) &= \{b \in [0, \infty) \mid \text{dist}(a, b) < \varepsilon\} \\
&= \left\{ b \in [0, \infty) \mid -\log \sup_{c \in [0, \infty)} \frac{c-a}{c-b} < \varepsilon \right\} \\
&= \left\{ b \in [0, \infty) \mid \begin{array}{l} -\log 1 = 0 < \varepsilon, \quad b < a \\ -\log \frac{a}{b} < \varepsilon, \quad b \geq a \end{array} \right\} \\
&= \{b \in [0, \infty) \mid b < ae^{\varepsilon}\}
\end{aligned}$$

$B(a, \varepsilon)$  is as in the picture:



if  $a = 0$ , then  $ae^{\varepsilon} = 0$ , thus  $B(0, \varepsilon) = \{b \in [0, \infty) \mid b \leq 0\} = \{0\}$ .

Remark. For the real line  $(-\infty, \infty)$ , since  $\lim_{c \rightarrow \pm\infty} \frac{c-a}{c-b} = 1$ ,

therefore,  $\text{dist}(a, b) = -\log \sup_{c \in (-\infty, \infty)} \frac{c-a}{c-b} = 0$ . Thus, any open ball

(contravariant, covariant) in  $(-\infty, \infty)$  is equal to  $(-\infty, \infty)$ .

Definition 2. For  $r \in K$ , define  $\frac{r}{0} = \begin{cases} 1, & r = 0 \\ \infty, & r > 0 \end{cases}$ .

Lemma 1.  $f, g \in \mathcal{P}(M)(X, Y)$ , if  $f \neq g$ , then  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} < 1$ ,  
and  $\sup_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} > 1$ .

Proof.  $f \neq g$  implies that there exists  $x \in X$ ,  $B \in \mathcal{B}$ , such that  $f(x, B) \neq g(x, B)$ . Let us denote  $Y \setminus B = B^c$ , since  $f(x, -)$  and  $g(x, -)$  are probability measures,

then either  $f(x, B) < g(x, B)$ ,  $f(x, B^c) > g(x, B^c)$ ;  
or  $f(x, B) > g(x, B)$ ,  $f(x, B^c) < g(x, B^c)$ .

Thus, we have the results.

Lemma 2.  $f, g \in \mathcal{P}(M)(X, Y)$ , then

$$a). \quad \sup_{\substack{t \in [0, 1], \\ f = tg + (1-t)h}} t = \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)},$$

therefore,  $\text{dist}(f, g) = -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)}$ .

$$b). \quad -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = \sup_{\substack{x \in X \\ B \in \mathcal{B}}} \log \frac{g(x, B)}{f(x, B)}.$$

c). View  $f, g$  as functions  $X \longrightarrow \mathcal{P}(Y)$ , then

$$\sup_{x \in X} \mathcal{P}(Y)(f(x, -), g(x, -)) = \text{dist}(f, g).$$

Proof. First we show that  $\inf_{g(x, B) \neq 0} \frac{f(x, B)}{g(x, B)} = \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)}$ .

If for some  $x' \in X$ ,  $B' \in \mathcal{B}$ ,  $g(x', B') = 0$ , we have two cases:

(1).  $f(x', B') = 0$ , then  $\frac{f(x', B')}{g(x', B')} = \frac{0}{0} = 1$ ,

(2).  $f(x', B') \neq 0$ , then  $\frac{f(x', B')}{g(x', B')} = \infty$ .

If  $f \neq g$ , from Lemma 1,  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} < 1$ . Therefore, those

$(x', B')$ 's (such that  $g(x', B') = 0$ ) do not effect the result of  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)}$ . Thus,  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = \inf_{g(x, B) \neq 0} \frac{f(x, B)}{g(x, B)}$ .

If  $f = g$ , then  $f(x, B) = g(x, B)$  for all  $x \in X, B \in \mathcal{B}$ , therefore

$$\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = 1 = \inf_{g(x, B) \neq 0} \frac{f(x, B)}{g(x, B)}.$$

a). Let  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = s_0$ . For any  $x \in X, B \in \mathcal{B}$ , we have

$$\frac{f(x, B)}{g(x, B)} \geq \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = s_0. \text{ So } f(x, B) \geq s_0 g(x, B) \text{ or}$$

$$f(x, B) - s_0 g(x, B) \geq 0.$$

$$\text{Define } h(x, B) = \frac{f(x, B) - s_0 g(x, B)}{1 - s_0},$$

(1). for each  $B \in \mathcal{B}$ ,  $h(\cdot, B)$  is  $\mathcal{A}$ -measurable; since  $f(\cdot, B)$  and  $g(\cdot, B)$  are  $\mathcal{A}$ -measurable;

(2).  $h(x, B) \geq 0$ : since  $f(x, B) - s_0 g(x, B) \geq 0$  and  $1 - s_0 \geq 0$ ;

(3).  $h(x, B) \leq 1$ : if let  $B^c = Y \setminus B$ , then  $h(x, B) \geq 0$ , and  $h(x, B^c) \geq 0$ ,

$$\text{and } h(x, B) + h(x, B^c) = \frac{1}{1 - s_0} (f(x, B) - s_0 g(x, B) + f(x, B^c) - s_0 g(x, B^c))$$

$$= \frac{1}{1 - s_0} ((f(x, B) + f(x, B^c)) - s_0 (g(x, B) + g(x, B^c)))$$

$$= \frac{1}{1 - s_0} (1 - s_0) = 1, \text{ this shows that } h(x, B) \leq 1.$$

$h(x, \cdot)$  is a measure, since  $f(x, \cdot)$  and  $g(x, \cdot)$  are. Thus,  $h \in \mathcal{P}(m)(X, Y)$ .

Let  $\sup_{\substack{t \in [0,1], \\ f = tg + (1-t)h}} t = t_0$ . Since  $f = s_0 g + (1-s_0)h$

thus,  $s_0 \leq t_0$ .

On the other hand, if  $f = tg + (1-t)h$ , i.e. for any  $x \in X, B \in \mathcal{B}$ ,  
 $f(x, B) = tg(x, B) + (1-t)h(x, B)$ .

If  $g(x, B) \neq 0$ , then  $\frac{f(x, B)}{g(x, B)} = t + (1-t)\frac{h(x, B)}{g(x, B)}$ , this implies

that  $\frac{f(x, B)}{g(x, B)} \geq t$ , therefore,  $\inf_{g(x, B) \neq 0} \frac{f(x, B)}{g(x, B)} \geq t$ , and

$$s_0 = \inf_{g(x, B) \neq 0} \frac{f(x, B)}{g(x, B)} \geq \sup_{\substack{t \in [0,1], \\ f = tg + (1-t)h}} t = t_0. \text{ Thus, } s_0 = t_0.$$

b). If  $f = g$ , the equality  $-\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = \sup_{\substack{x \in X \\ B \in \mathcal{B}}} \log \frac{g(x, B)}{f(x, B)}$

holds.

Now  $f \neq g$ :

1). if  $(x', B') \in A_1 = \{(x, B) \mid x \in X, B \in \mathcal{B}, f(x, B) = 0 \text{ and } g(x, B) \neq 0\}$ ,

then  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x', B')}{g(x', B')} = 0$ , so  $-\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = \infty$ ; and

$\sup_{\substack{x \in X \\ B \in \mathcal{B}}} \log \frac{g(x, B)}{f(x, B)} = \infty$ . Thus, b) holds.

2). if  $(x', B') \in A_2 = \{(x, B) \mid x \in X, B \in \mathcal{B}, f(x, B) \neq 0 \text{ and } g(x, B) = 0\}$ ,

then  $\frac{f(x', B')}{g(x', B')} = \infty$ , by Lemma 1 it does not effect the value of

$\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)}$ , therefore does not effect the value of  $-\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)}$ ;

and  $\frac{g(x', B')}{f(x', B')} = 0$ , so it does not effect the value of  $\sup_{\substack{x \in X \\ B \in \mathcal{B}}} \log \frac{g(x, B)}{f(x, B)}$ .

Let  $A = X \times \mathcal{B} \setminus A_1 \cup A_2$ , then any  $(x, B) \in A$ ,  $0 < \frac{f(x, B)}{g(x, B)}, \frac{g(x, B)}{f(x, B)} < \infty$

(in Definition 1, we defined  $\frac{0}{0} = 1$ ). Thus, we have

$$\begin{aligned} -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} &= -\log \inf_{(x, B) \in A} \frac{f(x, B)}{g(x, B)} = \sup_{(x, B) \in A} -\log \frac{f(x, B)}{g(x, B)} \\ &= \sup_{(x, B) \in A} \log \frac{g(x, B)}{f(x, B)} = \sup_{\substack{x \in X \\ B \in \mathcal{B}}} \log \frac{g(x, B)}{f(x, B)}. \end{aligned}$$

$$\begin{aligned} \text{c). } \sup_{x \in X} \rho(Y)(f(x, -), g(x, -)) &= \sup_{x \in X} \text{dist}(f(x, -), g(x, -)) \\ &= \sup_{x \in X} \left( -\log \sup_{\substack{s \in [0, 1] \\ \exists p \in \rho(Y) \\ f(x, -) = sg(x, -) + (1-s)p}} s \right) \end{aligned}$$

$$\begin{aligned} \text{but } \text{dist}(f, g) &= -\log \sup_{\substack{\exists t \in [0, 1], \exists h \in \rho(m)(X, Y) \\ f = tg + (1-t)h}} t = -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} \quad (\text{by a.}) \\ &= \sup_{\substack{x \in X \\ B \in \mathcal{B}}} \log \frac{g(x, B)}{f(x, B)} \quad (\text{by b.}) \\ &= \sup_{x \in X} \sup_{B \in \mathcal{B}} \log \frac{g(x, B)}{f(x, B)} \end{aligned}$$

$$\text{Thus, we have to show that } -\log \sup_{\substack{\exists s \in [0, 1], p \in \rho(Y) \\ f(x, -) = sg(x, -) + (1-s)p}} s = \sup_{B \in \mathcal{B}} \log \frac{g(x, B)}{f(x, B)}$$

But this is just the cases a). and b). again, for  $f(x, -)$ ,  $g(x, -)$  in  $\rho(m)(1, Y)$ .

Remark. From the proof of a). we know that if  $\text{dist}(f, g) = \varepsilon$ , then there is  $h \in \rho(m)(X, Y)$ , such that  $f = e^{-\varepsilon}g + (1 - e^{-\varepsilon})h$ .

Lemma 3.  $f, g \in \rho(m)(X, Y)$ , then  $\text{dist}(f, g) = 0$  iff  $f = g$ .

Proof. If  $f = g$ , then  $\text{dist}(f, g) = 0$ .

If  $\text{dist}(f, g) = 0$ , i.e.  $-\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = 0$  or  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = 1$ .

Assume  $\exists x \in X, B \in \mathcal{B}$ , such that  $f(x, B) > g(x, B)$ , then

$f(x, B^c) < g(x, B^c)$ , where  $B^c = Y \setminus B$ . This implies  $\frac{f(x, B^c)}{g(x, B^c)} < 1$ .

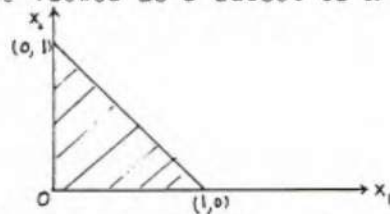
It contradicts the fact  $\inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{g(x, B)} = 1$ , therefore  $f = g$ .

Example 3.  $\mathcal{P}(m)(1, 3)$ .

Note that,  $\mathcal{P}(m)(1, 2) = [0, 1]$  as convex set.

$\mathcal{P}(m)(1, 3)$  is the free convex set generated by three elements.

$\mathcal{P}(m)(1, 3)$  could be viewed as a subset of  $\mathbb{R}^2$ :



Any  $b \in \mathcal{P}(m)(1, 3)$ ,  $b = (b_1, b_2, b_3)$ ,  $b_i \geq 0$  and  $\sum_{i=1}^3 b_i = 1$ .

We take  $b_1, b_2$  as coordinates of  $b$  on  $x_1$ -axis and  $x_2$ -axis.

Let  $a \in \mathcal{P}(m)(1, 3)$ ,  $\varepsilon > 0$ , the contravariant open ball with center  $a$  and radius  $\varepsilon$  is

$$\begin{aligned} B(\varepsilon, a) &= \{b \in \mathcal{P}(m)(1, 3) \mid \text{dist}(b, a) < \varepsilon\} \\ &= \{b \in \mathcal{P}(m)(1, 3) \mid -\log \min_i \frac{b_i}{a_i} < \varepsilon\} \\ &= \{b \in \mathcal{P}(m)(1, 3) \mid \min_i \frac{b_i}{a_i} > e^{-\varepsilon}\} \end{aligned}$$

For  $b \in B(\varepsilon, a)$ ,  $\frac{b_i}{a_i} \geq \min_i \frac{b_i}{a_i} > e^{-\varepsilon}$ ,

if  $a_i \neq 0$ ,  $b_i > a_i e^{-\epsilon}$ ,

if  $a_i = 0$ ,  $b_i \geq 0 = a_i = a_i e^{-\epsilon}$ .

From  $b_3 > a_3 e^{-\epsilon}$ , or  $1 - (b_1 + b_2) > a_3 e^{-\epsilon}$ , we get  $b_1 + b_2 < 1 - a_3 e^{-\epsilon}$ .

Note that,  $a_1 > a_1 e^{-\epsilon}$ ,  $a_2 > a_2 e^{-\epsilon}$ , and

$$\begin{aligned} (1 - a_3 e^{-\epsilon}) - (a_1 + a_2) &= 1 - (a_1 + a_2) - a_3 e^{-\epsilon} \\ &= a_3 - a_3 e^{-\epsilon} \\ &= a_3(1 - e^{-\epsilon}) \geq 0 \end{aligned}$$

so,  $a_1 + a_2 \leq 1 - a_3 e^{-\epsilon}$ .

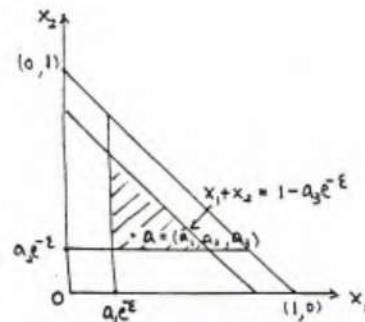
Thus,  $F(\epsilon, a)$  is the region bounded by

$$\begin{cases} x_1 \geq a_1 e^{-\epsilon} \\ x_2 \geq a_2 e^{-\epsilon} \\ x_1 + x_2 \leq 1 - a_3 e^{-\epsilon} \end{cases}$$

("=" holds, only when  $a_i = 0$ ,  $i = 1, 2, 3$ )

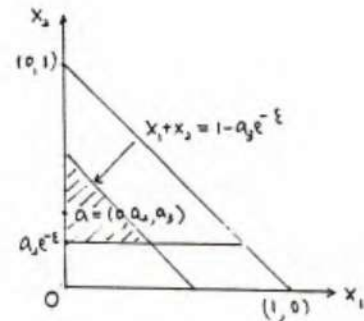
(1).  $a_i \neq 0$ ,  $i=1, 2, 3$ . then

$$B(\epsilon, a) = \begin{cases} x_1 > a_1 e^{-\epsilon} \\ x_2 > a_2 e^{-\epsilon} \\ x_1 + x_2 < 1 - a_3 e^{-\epsilon} \end{cases}$$



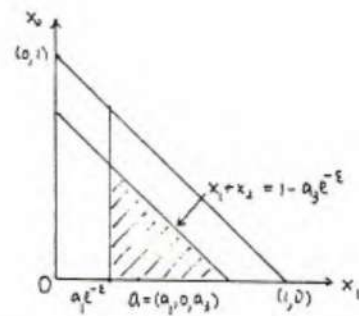
(2).  $a_1 = 0, a_2 \neq 0, a_3 \neq 0$ , then

$$B(\varepsilon, a) = \begin{cases} x_1 = 0 \\ x_2 > a_2 e^{-\varepsilon} \\ x_1 + x_2 < 1 - a_3 e^{-\varepsilon} \end{cases}$$



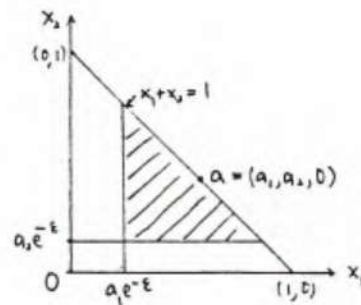
(3).  $a_1 \neq 0, a_2 = 0, a_3 \neq 0$ , then

$$B(\varepsilon, a) = \begin{cases} x_1 > a_1 e^{-\varepsilon} \\ x_2 = 0 \\ x_1 + x_2 < 1 - a_3 e^{-\varepsilon} \end{cases}$$



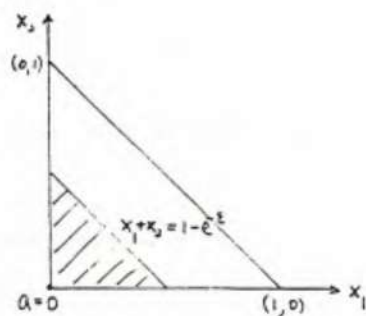
(4).  $a_1 \neq 0, a_2 \neq 0, a_3 = 0$ , then

$$B(\varepsilon, a) = \begin{cases} x_1 > a_1 e^{-\varepsilon} \\ x_2 > a_2 e^{-\varepsilon} \\ x_1 + x_2 = 1 \end{cases}$$



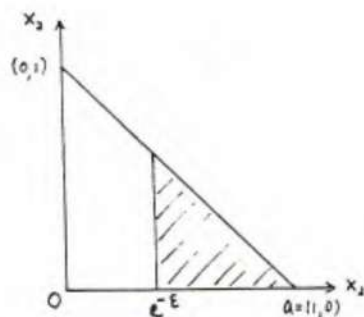
(5).  $a_1 = 0, a_2 = 0, a_3 = 1$ , then

$$B(\varepsilon, a) = \begin{cases} x_1 = 0 \\ x_2 = 0 \\ x_1 + x_2 < 1 - e^{-\varepsilon} \end{cases}$$



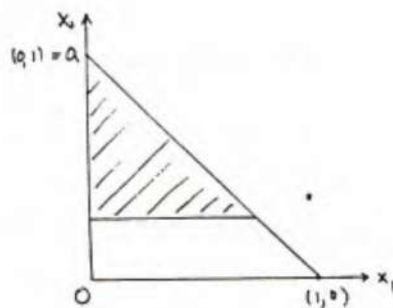
(6).  $a_1 = 1, a_2 = 0, a_3 = 0$ , then

$$B(\varepsilon, a) = \begin{cases} x_1 > e^{-\varepsilon} \\ x_2 = 0 \\ x_1 + x_2 = 1 \end{cases}$$



(7).  $a_1 = 0, a_2 = 1, a_3 = 0$ , then

$$B(\varepsilon, a) = \begin{cases} x_1 = 0 \\ x_2 > e^{-\varepsilon} \\ x_1 + x_2 = 1 \end{cases}$$



Example 4.  $\mathcal{P}(m)(1,n)$ ,  $n < \infty$ .

at  $\mathcal{P}(m)(1,n)$ , then  $a = (a_1, \dots, a_n)$ ,  $a_i \geq 0$ , and  $\sum_{i=1}^n a_i = 1$ .

We can view  $\mathcal{P}(m)(1,n)$  as a subset of  $R^{n-1}$ , such that at  $\mathcal{P}(m)(1,n)$

$a = (a_1, \dots, a_{n-1})$ ,  $a_i \geq 0$ , and  $\sum_{i=1}^{n-1} a_i \leq 1$ .

The contravariant open ball with center  $a$  and radius  $\varepsilon$  is

$$\begin{aligned} B(\varepsilon, a) &= \{ b \in \mathcal{P}(m)(1,n) \mid \text{dist}(b, a) < \varepsilon \} \\ &= \{ b \in \mathcal{P}(m)(1,n) \mid -\log \min_i \frac{b_i}{a_i} < \varepsilon, \text{ or } \min_i \frac{b_i}{a_i} > e^{-\varepsilon} \} \end{aligned}$$

$b \in B(\varepsilon, a)$ , if  $a_i \neq 0$ , then  $\frac{b_i}{a_i} \geq \min_i \frac{b_i}{a_i} > e^{-\varepsilon}$ , or  $b_i > a_i e^{-\varepsilon}$ ;

if  $a_i = 0$ , then  $b_i \geq 0 = a_i = a_i e^{-\varepsilon}$ .

$B(\varepsilon, a)$  is the region  $\left\{ \begin{array}{l} x_1 \geq a_1 e^{-\varepsilon} \\ \vdots \\ x_{n-1} \geq a_{n-1} e^{-\varepsilon} \\ x_1 + x_2 + \dots + x_{n-1} \leq 1 - (1 - \sum_{i=1}^{n-1} a_i) e^{-\varepsilon} = 1 - a_n e^{-\varepsilon} \end{array} \right.$   
 ("=" holds only when  $a_i = 0, i=1, \dots, n$ )

The Euclidean distance from  $a$  to  $x_i = a_i e^{-\varepsilon}$ ,  $i = 1, \dots, n-1$ ,

is  $a_i - a_i e^{-\varepsilon} = a_i(1 - e^{-\varepsilon})$ ; the Euclidean distance from  $a$  to

$$\sum_{i=1}^{n-1} x_i = 1 - a_n e^{-\varepsilon} \quad \text{is} \quad \frac{|a_1 + \dots + a_{n-1} - (1 - a_n e^{-\varepsilon})|}{\sqrt{n-1}}$$

$= \frac{1}{\sqrt{n-1}} (1 - e^{-\varepsilon}) a_n$  (the Euclidean distance from  $a = (a_1, \dots, a_{n-1})$ )

to  $\sum_{i=1}^{n-1} b_i x_i = c$  is  $\frac{|\sum_{i=1}^{n-1} b_i a_i - c|}{\sqrt{\sum_{i=1}^{n-1} a_i^2}}$ .

Corollary. The contravariant topology on  $\mathcal{P}(m)(1,n)$  is

equivalent to the usual topology on  $\mathcal{P}(m)(1,n)$  restricted from  $\mathbb{R}^{n-1}$ .

Proof. Let us denote  $U(\delta, a)$  as the intersection of  $\mathcal{P}(m)(1,n)$  and the usual open ball of center  $a$  and radius  $\delta$  in  $\mathbb{R}^{n-1}$ .

If we have  $B(\varepsilon, a)$ , let  $\delta = \min\left\{a_1(1-e^{-\varepsilon}), \frac{1}{\sqrt{n-1}}(1-e^{-\varepsilon})a_n\right\}$

then  $U(\delta, a) \subseteq B(\varepsilon, a)$ .

Conversely, if we have  $U(\delta, a)$ ,

in the case  $n=3$ ,  $B(\varepsilon, a)$  is a

triangle with vertices

$$(a_1 e^{-\varepsilon}, 1 - (1 - a_2) e^{-\varepsilon}),$$

$$(a_1 e^{-\varepsilon}, a_2 e^{-\varepsilon}) \text{ and}$$

$$(1 - (1 - a_1) e^{-\varepsilon}, a_2 e^{-\varepsilon}). \text{ The Euclidean}$$

distance between  $a$  and

$$(a_1 e^{-\varepsilon}, 1 - (1 - a_2) e^{-\varepsilon}) \text{ is } s_1 =$$

$$\begin{aligned} & \sqrt{a_1^2(1-e^{-\varepsilon})^2 + (a_2 - (1-a_2)e^{-\varepsilon})^2} \\ & = (1-e^{-\varepsilon})\sqrt{a_1^2 + (1-a_2)^2}; \end{aligned}$$

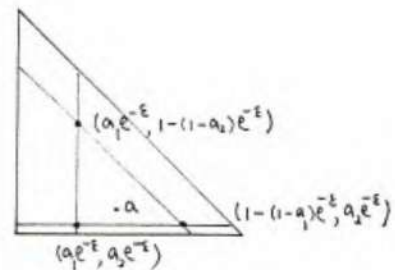
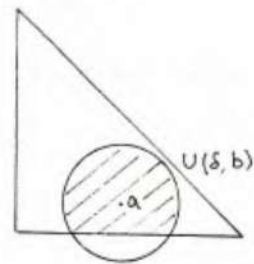
The Euclidean distance between  $a$

$$\text{and } (a_1 e^{-\varepsilon}, a_2 e^{-\varepsilon}) \text{ is } s_2 =$$

$$\sqrt{a_1^2(1-e^{-\varepsilon})^2 + a_2^2(1-e^{-\varepsilon})^2} =$$

$$(1-e^{-\varepsilon})\sqrt{a_1^2 + a_2^2}; \text{ the distance between } a \text{ and } (1 - (1 - a_1) e^{-\varepsilon}, a_2 e^{-\varepsilon})$$

$$\text{is } s_3 = \sqrt{(a_1 - (1 - (1 - a_1) e^{-\varepsilon}))^2 + a_2^2(1-e^{-\varepsilon})^2} = (1-e^{-\varepsilon})\sqrt{(1-a_1)^2 + a_2^2}.$$



If we choose  $\varepsilon > 0$ , such that  $s_i < \min\{\delta, a_i \frac{a_j}{|a_j|}\}$  for  $a_j \neq 0$ ;  
 $s_i < \min\{\delta, a_i\}$  for  $a_j = 0$ , then  $B(\varepsilon, a) \subset U(\delta, a)$ . Similarly, we  
 can show  $B(\varepsilon, a) \subset U(\delta, a)$ , for  $n > 3$ .

Now, let us look at the covariant topology on  $\mathcal{P}(m)(1, n)$ .

The covariant open ball with center  $a$  and radius  $\varepsilon$  is

$$B(a, \varepsilon) = \{b \in \mathcal{P}(m)(1, n) \mid \text{dist}(a, b) < \varepsilon\}$$

$$= \{b \in \mathcal{P}(m)(1, n) \mid -\log \min_i \frac{a_i}{b_i} < \varepsilon \text{ or } \min_i \frac{a_i}{b_i} > e^{-\varepsilon}\}$$

Note that, if  $b \in B(a, \varepsilon)$ , then  $a_i = 0$  implies  $b_i = 0$ , because  
 if  $a_i = 0$ ,  $b_i \neq 0$ , then  $-\log \min_j \frac{a_j}{b_j} = -\log 0 = \infty$ .

For  $b \in B(a, \varepsilon)$ ,  $\frac{a_i}{b_i} \geq \min_i \frac{a_i}{b_i} > e^{-\varepsilon}$ , that is  $b_i \leq a_i e^\varepsilon$  ("=" holds

only when  $a_i = 0$ ). From  $b_n \leq a_n e^\varepsilon$ , or  $1 - \sum_{i=1}^{n-1} b_i \leq (1 - \sum_{i=1}^{n-1} a_i) e^\varepsilon$ ,

we have  $\sum_{i=1}^{n-1} b_i \geq 1 - (1 - \sum_{i=1}^{n-1} a_i) e^\varepsilon = 1 - a_n e^\varepsilon$ .

Therefore,  $B(a, \varepsilon)$  is the  
 region ("=" holds only when  
 $a_i = 0$  or  $a_n = 0$ )

$$\begin{cases} x_i \leq a_i e^\varepsilon, & i = 1, \dots, n-1 \\ \sum_{i=1}^{n-1} x_i \geq 1 - a_n e^\varepsilon \end{cases} \quad (*)$$

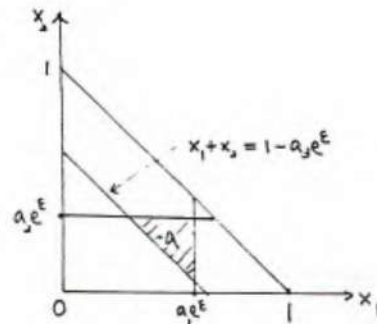
Note that,  $a$  is in this region, since  $a_i \leq a_i e^\varepsilon$ , and

$$\sum_{i=1}^{n-1} a_i = 1 - a_n \geq 1 - a_n e^\varepsilon.$$

For example,  $n = 3$ ,

(1).  $a_i \neq 0$ ,  $i = 1, 2, 3$ . Then

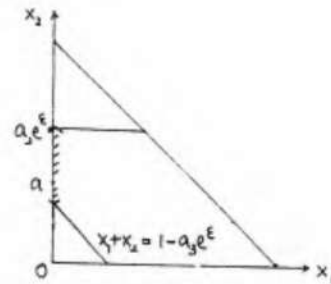
$$B(a, \varepsilon) = \begin{cases} x_1 < a_1 e^\varepsilon \\ x_2 < a_2 e^\varepsilon \\ x_1 + x_2 > 1 - a_3 e^\varepsilon \end{cases}$$



(2).  $a_1 = 0, a_2 \neq 0, a_3 \neq 0$

then

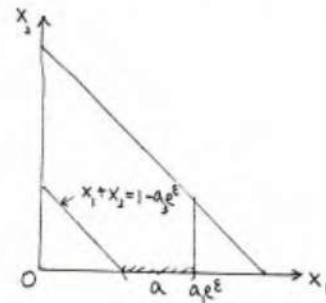
$$B(a, \varepsilon) = \begin{cases} x_1 = 0 \\ x_2 < a_2 e^\varepsilon \\ x_1 + x_2 > 1 - a_3 e^\varepsilon \end{cases}$$



(3).  $a_1 \neq 0, a_2 = 0, a_3 \neq 0$

then

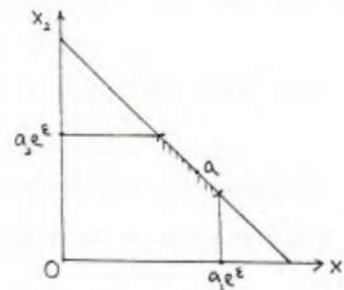
$$B(a, \varepsilon) = \begin{cases} x_1 < a_1 e^\varepsilon \\ x_2 = 0 \\ x_1 + x_2 > 1 - a_3 e^\varepsilon \end{cases}$$



(4).  $a_1 \neq 0, a_2 \neq 0, a_3 = 0$

then

$$B(a, \varepsilon) = \begin{cases} x_1 < a_1 e^\varepsilon \\ x_2 < a_2 e^\varepsilon \\ x_1 + x_2 = 1 \end{cases}$$

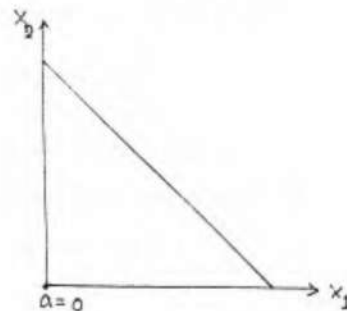


(5).  $a_1 = 0, a_2 = 0, a_3 = 1,$

i.e.  $a$  is the origin, then

$$B(a, \varepsilon) = \begin{cases} x_1 = 0 \\ x_2 = 0 \\ x_1 + x_2 > 1 - e^\varepsilon \end{cases} = \{a\}$$

note that,  $1 - e^\varepsilon < 0.$



$$(6). a_1 = 1, a_2 = 0, a_3 = 0,$$

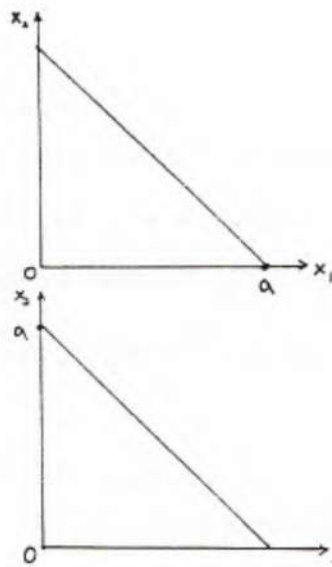
then

$$B(a, \varepsilon) = \begin{cases} x_1 < e^\varepsilon \\ x_2 = 0 \\ x_1 + x_2 = 1 \end{cases} = \{a\}$$

$$(7). a_1 = 0, a_2 = 1, a_3 = 0,$$

then

$$B(a, \varepsilon) = \begin{cases} x_1 = 0 \\ x_2 < e^\varepsilon \\ x_1 + x_2 = 1 \end{cases} = \{a\}$$



We see that the covariant topology on  $\mathcal{P}(m)(1,3)$  is the disjoint union of the usual topology on each open face.

From the description of  $B(a, \varepsilon)$ , (\*) in p.2.28, we have the general result,

Proposition 2. The covariant topology on  $\mathcal{P}(m)(1,n)$  (simplex) is the disjoint union of the usual topology on each open face.

Now we go back to the end of chapter 1, show that the closed functor  $\pi_0: \mathcal{U} \longrightarrow \mathcal{P}$  induces "face structure" on  $\mathcal{P}(m)(1,n)$ .

Recall that  $\pi_0(r) = \begin{cases} 1, & r < \infty \\ 0, & r = \infty \end{cases}$ , for  $r \in [0, \infty]$ .  $\pi_0$  induces a poset structure on  $\mathcal{P}(m)(1,n)$ , such that  $a, b \in \mathcal{P}(m)(1,n)$ ,  $a \leq b$  iff  $\text{dist}(a, b) < \infty$ . We define an equivalence relation on  $\mathcal{P}(m)(1,n)$ :  $a \sim b$  iff  $a \leq b$  and  $b \leq a$ , i.e.  $a \sim b$  iff

$\text{dist}(a,b) + \text{dist}(b,a) < \infty$ .

Lemma.  $a \sim b$  iff  $a, b$  lie in the same face.

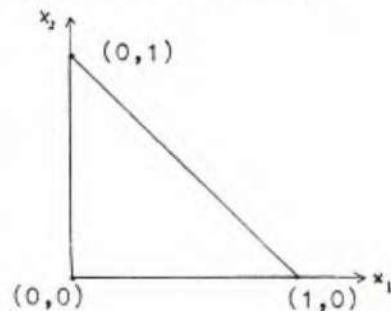
Proof.  $a = (a_i), b = (b_i), i = 1, \dots, n$ .  $a, b$  lie in the same face iff for any  $i$ ,  $a_i = 0$  implies  $b_i = 0$ , and  $b_i = 0$  implies  $a_i = 0$ . Since  $\text{dist}(a,b) = -\log \min_i \frac{a_i}{b_i}$  and we defined  $\frac{0}{0} = 1$ , this is equivalent to  $\text{dist}(b,a) < \infty$  and  $\text{dist}(a,b) < \infty$ , this is  $a \sim b$ .

Now we define an order relation  $\leq'$  on  $\mathcal{P}(m)(1,n)/\sim$ ,  $\bar{a}, \bar{b} \in \mathcal{P}(m)(1,n)/\sim$ ,  $\bar{a} \leq' \bar{a}$ ; for  $\bar{a} \neq \bar{b}$ ,  $\bar{a} \leq' \bar{b}$  iff  $\text{dist}(a,b) = \infty$  and  $\text{dist}(b,a) < \infty$ . This is well defined: if  $a \sim a', b \sim b'$ , then  $\text{dist}(b',a') \leq \text{dist}(b',b) + \text{dist}(b,a) + \text{dist}(a,a') < \infty$ ; if  $\text{dist}(a',b') < \infty$ , then  $\text{dist}(a,b) \leq \text{dist}(a,a') + \text{dist}(a',b') + \text{dist}(b',b) < \infty$ , this is a contradiction, thus  $\text{dist}(a',b') = \infty$ .

If  $\bar{a} \leq' \bar{b}$ ,  $\bar{b} \leq' \bar{c}$ , i.e.  $\text{dist}(a,b) = \infty = \text{dist}(b,c)$ ,  $\text{dist}(b,a) < \infty$  and  $\text{dist}(c,b) < \infty$ . Then  $\text{dist}(c,a) \leq \text{dist}(c,b) + \text{dist}(b,a) < \infty$ ; if  $\text{dist}(a,c) < \infty$ , then  $\text{dist}(a,b) \leq \text{dist}(a,c) + \text{dist}(c,b) < \infty$ , this is a contradiction, therefore  $\text{dist}(a,c) = \infty$ , i.e.  $\bar{a} \leq' \bar{c}$ .

Conclusion.  $(\mathcal{P}(m)(1,n)/\sim, \leq')$  gives the poset of faces of  $\mathcal{P}(m)(1,n)$  with the usual inclusion.

Example.  $\mathcal{P}(m)(1,3)$ . We denote  
 01 for the face determined  
 by  $(0,1), (0,0)$ ;  
 10 for the face determined  
 by  $(0,0), (1,0)$ ;



11 for the face determined by  $(0,1)$ ,  $(1,0)$ ;

012 for the two dimensional face.

Note that,  $(0,1)_1 = 0$ ,  $(0,1)_2 = 1$ ,  $(0,1)_3 = 0$ . If  $b$  is in 01,

$b \neq (0,1)$ , then  $b_1 = 0$ ,  $b_2 \neq 0$ ,  $b_3 \neq 0$ .  $\text{dist}((0,1), b) =$

$$-\log \min_i \frac{(0,1)_i}{b_i} = -\log \frac{0}{b_3} = \infty, \text{ but } \text{dist}(b, (0,1)) = -\log \min_i \frac{b_i}{(0,1)_i}$$

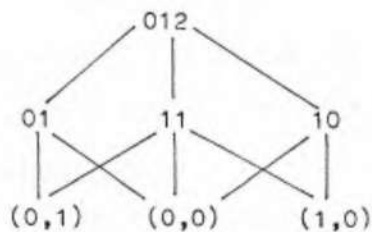
$$= -\log b_2 < \infty, \text{ because } \frac{r}{0} = \infty, \text{ for } r \neq 0.$$

It is easy to see that if  $a$  is in 01,  $b$  is in 10, then  $\text{dist}(a, b)$

$= \text{dist}(b, a) = \infty$ , where  $a \neq (0,0)$  and  $b \neq (0,0)$ . Also, if  $a$  is in 01,  $b$  is in 012, then  $\text{dist}(a, b) = -\log \min_i \frac{a_i}{b_i} = -\log 0 = \infty$ ,

but  $\text{dist}(b, a) = -\log \min_i \frac{b_i}{a_i} < \infty$ .

Thus, we have the following order relation on  $\beta(m)(1,3)/\sim$



From the above discussion we see that the contravariant topology on  $\mathcal{P}(M)(1,n)$  is Hausdorff and compact.

Proposition 3. For any pair of measurable spaces  $(X,\mathcal{A}), (Y,\mathcal{B})$ , the contravariant topology of  $\mathcal{P}(M)(X,Y)$  is Hausdorff.

We need one lemma from chapter 1 §4.

Lemma.  $C$  is a convex set,  $a,b,c \in C$ . If  $a = tb + (1-t)c$  and  $0 < s < t < 1$ , then for some  $c' \in C$ ,  $a = sb + (1-s)c'$ .

Remark.  $a,b \in C$ , if  $c \in B(\delta,a) \cap B(\delta,b)$ , then  $\exists s \in [0,1], s > e^{-\delta}$ , and  $\exists a', b' \in C$ , such that  $c = sa + (1-s)a'$  and  $c = sb + (1-s)b'$ .  
Say  $\text{dist}(c,a) = \varepsilon_1$ ,  $\text{dist}(c,b) = \varepsilon_2$ , and  $\varepsilon_1 < \varepsilon_2$ , or

$$e^{-\varepsilon_1} > e^{-\varepsilon_2} > e^{-\delta}. \text{ Since } \varepsilon_2 = -\log \sup_{\substack{s \in [0,1], \exists b' \in C \\ c = sb + (1-s)b'}} s, \text{ or } e^{-\varepsilon_2} = \sup_{\substack{s \in [0,1], \exists b' \in C \\ c = sb + (1-s)b'}} s$$

so,  $\exists s, \exists b' \in C$ , such that  $e^{-\delta} < s < e^{-\varepsilon_2}$  and  $c = sb + (1-s)b'$ .

Since  $e^{-\varepsilon_2} < e^{-\varepsilon_1} = \sup_{\substack{t \in [0,1], \exists a' \in C \\ c = ta + (1-t)a'}} t$ , so  $\exists t, \exists a'' \in C$ , such that

$e^{-\varepsilon_2} < t < e^{-\varepsilon_1}$ , and  $c = ta + (1-t)a''$ . Using the above Lemma, for  $s < e^{-\varepsilon_2} < t$ , we get  $a' \in C$ , and  $c = sa + (1-s)a'$ .

Proof of the proposition:

If  $f, g \in \mathcal{P}(M)(X,Y)$  and  $f \neq g$ , then there exists  $x' \in X, B' \in \mathcal{B}$ , such that  $f(x', B') \neq g(x', B')$ . Let  $e(x, B) = \max\{f(x, B), g(x, B)\}$ ,

where  $x \in X, B \in \mathcal{B}$ ; and  $t = \frac{1}{e(x', B') + e(x', B'^c)}$ , where  $B^c = Y \setminus B$ .

Note that,  $1 < e(x', B') + e(x', B'^c) \leq 2$ , therefore  $\frac{1}{2} \leq t < 1$ .

Let  $\delta = -\log t$  or  $e^{-\delta} = t$ . If  $h \in B(\delta, f) \cap B(\delta, g)$ , from the

remark, there exists  $s > e^{-\delta}$ ,  $f', g' \in \mathcal{P}(M)(X, Y)$ , such that  
 $h = sf + (1-s)f'$  and  $h = sg + (1-s)g'$ , i.e. for any  $x \in X$ ,  $B \in \mathcal{B}$ ,

$$\text{we have } h(x, B) = sf(x, B) + (1-s)f'(x, B)$$

$$h(x, B) = sg(x, B) + (1-s)g'(x, B)$$

then  $h(x, B) \geq sf(x, B)$  and  $h(x, B) \geq g(x, B)$ , therefore,

$$h(x, B) \geq se(x, B).$$

Specially, for  $x = x'$ ,  $B = B'$  and  $B = B'^c$ ,

$$1 = h(x', B') + h(x', B'^c) \geq s(e(x', B') + e(x', B'^c)), \text{ this implies}$$

$$s \leq \frac{1}{e(x', B') + e(x', B'^c)} = t = e^{-\delta}. \text{ This is a contradiction,}$$

thus  $E(\delta, f) \cap E(\delta, g) = \emptyset$ .

Remark. In case of covariant topology of  $\mathcal{P}(M)(X, Y)$ , we may consider  $d(x, B) = \min\{f(x, B), g(x, B)\}$ , and let  $t = d(x', B') + d(x', B'^c)$ ,  $\delta = -\log t$ , then we could get the conclusion: the covariant topology of  $\mathcal{P}(M)(X, Y)$  is Hausdorff.

Now we are going to discuss the completeness of  $\mathcal{P}(M)(X, Y)$ .

Definition 3.  $f_n \in \mathcal{P}(M)(X, Y)$ ,  $n = 1, 2, \dots$  is a Cauchy sequence, if for any  $\varepsilon > 0$ ,  $\exists N(\varepsilon)$ , such that  $\text{dist}(f_k, f_p) < \varepsilon$ , for any  $k, p > N(\varepsilon)$ . We denote a sequence  $f_n$ ,  $n = 1, 2, \dots$  by  $\{f_n\}$ .

Theorem.  $\mathcal{P}(M)(X, Y)$  is sequentially Cauchy complete, i.e. if  $f_n$ ,  $n = 1, 2, \dots$  is a Cauchy sequence in  $\mathcal{P}(M)(X, Y)$ , then there exists  $f \in \mathcal{P}(M)(X, Y)$ , such that  $\text{dist}(f, f_p) \rightarrow 0$  as  $p \rightarrow \infty$  and  $\text{dist}(f_k, f) \rightarrow 0$  as  $k \rightarrow \infty$ .

Proof. First note, if  $\{f_n\}$  is a Cauchy sequence in  $\mathcal{P}(m)(X, Y)$ , then for  $x \in X$ ,  $B \in \mathcal{B}$ ,  $\{f_n(x, B)\}$  is a Cauchy sequence in  $\mathbb{R}$  (the reals) in the usual sense, because for any  $\varepsilon > 0$ ,  $\exists N(\varepsilon)$ , as  $k, p > N(\varepsilon)$ ,  $\text{dist}(f_k, f_p) < \varepsilon$ , then

$$|f_k(x, B) - f_p(x, B)| \leq \sup_{\substack{x \in X \\ B \in \mathcal{B}}} |f_k(x, B) - f_p(x, B)| < \text{dist}(f_k, f_p) < \varepsilon$$

the second inequality holds, because the proposition 6, in last section.

Since for any  $x \in X$ ,  $B \in \mathcal{B}$ , the Cauchy sequence  $\{f_n(x, B)\}$  has a limit in  $\mathbb{R}$ , we define  $f(x, B) = \lim_{k \rightarrow \infty} f_k(x, B)$ . We have to check that  $f \in \mathcal{P}(m)(X, Y)$  and  $\text{dist}(f_k, f) \rightarrow 0$  as  $k \rightarrow \infty$ ;  $\text{dist}(f, f_p) \rightarrow 0$  as  $p \rightarrow \infty$ .

a). For each  $x \in X$ ,  $f(x, -)$  is a probability measure on  $(Y, \mathcal{B})$ :

$$f(x, Y) = \lim_{k \rightarrow \infty} f_k(x, Y) = 1.$$

Let  $B_i \in \mathcal{B}$ ,  $i = 1, 2, \dots$ ,  $B_i \cap B_j = \emptyset$ ,  $i \neq j$ ,

$$\begin{aligned} \text{then } f(x, \bigcup_{i=1}^{\infty} B_i) &= \lim_{k \rightarrow \infty} f_k(x, \bigcup_{i=1}^{\infty} B_i) \\ &= \lim_{k \rightarrow \infty} \sum_{i=1}^{\infty} f_k(x, B_i) \\ &= \sum_{i=1}^{\infty} \lim_{k \rightarrow \infty} f_k(x, B_i) \\ &= \sum_{i=1}^{\infty} f(x, B_i) \end{aligned}$$

the third equality holds, because from Theorem 8.39 (Tom Apostol, Mathematical Analysis): "assume that  $\lim_{p, q \rightarrow \infty} f(p, q) = a$ ; for each

fixed  $p$  assume that the limit  $\lim_{q \rightarrow \infty} f(p, q)$  exists. Then the limit

$\lim_{p \rightarrow \infty} (\lim_{q \rightarrow \infty} f(p, q))$  also exists and has the value  $a$ . Here we take

$$f(p, q) = f_q(x, \bigcup_{i=1}^p B_i), \text{ since } \lim_{q \rightarrow \infty} f(p, q) = \lim_{q \rightarrow \infty} f_q(x, \bigcup_{i=1}^p B_i) = f(x, \bigcup_{i=1}^p B_i)$$

$$\text{and } \lim_{q \rightarrow \infty} f_q(x, \bigcup_{i=1}^p B_i) = \lim_{q \rightarrow \infty} \sum_{i=1}^p f_q(x, B_i) = \sum_{i=1}^p f(x, B_i), \text{ thus}$$

$$\lim_{p \rightarrow \infty} (\lim_{q \rightarrow \infty} f(p, q)) = \lim_{p \rightarrow \infty} \sum_{i=1}^p f(x, B_i) = \sum_{i=1}^{\infty} f(x, B_i).$$

b). For each  $B \in \mathcal{B}$ ,  $f(-, B)$  is  $\mathcal{A}$ -measurable. This is because for each  $k$ ,  $f_k(-, B)$  is  $\mathcal{A}$ -measurable, so is the limit  $\lim_{k \rightarrow \infty} f_k(-, B) = f(-, B)$ .

c). We show that  $\text{dist}(f, f_p) \rightarrow 0$  as  $p \rightarrow \infty$  and  $\text{dist}(f_k, f) \rightarrow 0$  as  $k \rightarrow \infty$ .

For any  $\varepsilon > 0$ ,  $\exists N(\frac{\varepsilon}{2})$ , such that as  $p, k > N(\frac{\varepsilon}{2})$ ,  $\text{dist}(f_k, f_p) < \frac{\varepsilon}{2}$ ,

$$\text{i.e. } -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f_k(x, B)}{f_p(x, B)} < \frac{\varepsilon}{2}, \text{ or } \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f_k(x, B)}{f_p(x, B)} > e^{-\frac{\varepsilon}{2}}.$$

$$\text{So, for any } x \in X, B \in \mathcal{B}, \frac{f_k(x, B)}{f_p(x, B)} \geq \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f_k(x, B)}{f_p(x, B)} > e^{-\frac{\varepsilon}{2}},$$

$$\text{then } \lim_{k \rightarrow \infty} \frac{f_k(x, B)}{f_p(x, B)} = \frac{\lim_{k \rightarrow \infty} f_k(x, B)}{f_p(x, B)} = \frac{f(x, B)}{f_p(x, B)} \geq e^{-\frac{\varepsilon}{2}},$$

$$\text{and } \lim_{p \rightarrow \infty} \frac{f_k(x, B)}{f_p(x, B)} = \frac{f_k(x, B)}{\lim_{p \rightarrow \infty} f_p(x, B)} = \frac{f_k(x, B)}{f(x, B)} \geq e^{-\frac{\varepsilon}{2}}.$$

$$\text{But that is, } \text{dist}(f, f_p) = -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f(x, B)}{f_p(x, B)} \leq \frac{\epsilon}{2} < \epsilon,$$

$$\text{and } \text{dist}(f_k, f) = -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f_k(x, B)}{f(x, B)} \leq \frac{\epsilon}{2} < \epsilon.$$

This complete the proof of the theorem.

Since the generalized metric  $\text{dist}$  on  $\mathcal{P}(M)(X, Y)$  is not symmetric, the definition of Cauchy sequence above is rather strong. We can weaken the requirement for being a Cauchy sequence in the following way,

Definition 4.  $\{f_n\}$  is a sequence in  $\mathcal{P}(M)(X, Y)$ .

$\{f_n\}$  is a contravariant Cauchy sequence, if for any  $\epsilon > 0$ ,

$\exists N(\epsilon)$ , such that as  $p > k > N(\epsilon)$ ,  $\text{dist}(f_k, f_p) < \epsilon$  ;

$\{f_n\}$  is a covariant Cauchy sequence, if for any  $\epsilon > 0$ ,  $\exists N(\epsilon)$ ,

such that as  $p > k > N(\epsilon)$ ,  $\text{dist}(f_p, f_k) < \epsilon$  .

Example.  $f_n \in \mathcal{P}(M)(1, 2) = \mathcal{P}(2)$ ,  $f_n(0) = \frac{1}{2^n}$ ,  $f_n(1) = \frac{2^n - 1}{2^n}$ ,  $n=1, 2, \dots$

taking  $2 = \{0, 1\}$  .

$\{f_n\}$  is a contravariant Cauchy sequence, since

$$\begin{aligned} \text{dist}(f_n, f_{n+p}) &= -\log \min \left\{ \frac{f_n(0)}{f_{n+p}(0)}, \frac{f_n(1)}{f_{n+p}(1)} \right\} \\ &= -\log \min \left\{ \frac{2^{n+p}}{2^n}, \frac{2^{n+p}(2^n-1)}{2^n(2^{n+p}-1)} \right\} \\ &= -\log \frac{2^p(2^n-1)}{2^{n+p}-1} \end{aligned}$$

but  $\frac{2^p(2^n-1)}{2^{n+p-1}} > \frac{2^p(2^n-1)}{2^{n+p}} = \frac{2^n-1}{2^n}$ , then  $2^n > \frac{1}{1-e^{-\epsilon}}$  or  $\frac{2^n-1}{2^n} > e^{-\epsilon}$  implies  $-\log \frac{2^n-1}{2^n} < \epsilon$ . Thus, if we take  $N(\epsilon) > \log_2 \frac{1}{1-e^{-\epsilon}}$ , then for any  $n > N(\epsilon)$ ,  $-\log \frac{2^p(2^n-1)}{2^{n+p-1}} < -\log \frac{2^n-1}{2^n} < \epsilon$ ,

i.e. for any positive integer  $p$ ,  $\text{dist}(f_n, f_{n+p}) < \epsilon$ .

$\{f_n\}$  is not a covariant Cauchy sequence, since

$$\begin{aligned} \text{dist}(f_{n+p}, f_n) &= -\log \min \left\{ \frac{f_{n+p}(0)}{f_n(0)}, \frac{f_{n+p}(1)}{f_n(1)} \right\} \\ &= -\log \min \left\{ \frac{2^n}{2^{n+p}}, \frac{2^n(2^{n+p}-1)}{2^{n+p}(2^n-1)} \right\} \\ &= -\log \frac{2^n}{2^{n+p}} = -\log \frac{1}{2^p} \rightarrow \infty, \text{ as } p \rightarrow \infty. \end{aligned}$$

Proposition 4. If  $\{f_n\}$  is a contravariant Cauchy sequence in  $\mathcal{P}(M)(X, Y)$ , then there exists  $f \in \mathcal{P}(M)(X, Y)$ , such that  $\text{dist}(f_k, f) \rightarrow 0$ , as  $k \rightarrow \infty$ ;

if  $\{f_n\}$  is a covariant Cauchy sequence in  $\mathcal{P}(M)(X, Y)$ , then there exists  $f \in \mathcal{P}(M)(X, Y)$ , such that  $\text{dist}(f, f_k) \rightarrow 0$ , as  $k \rightarrow \infty$ .

Proof. The construction of  $f$  and the proof of  $f \in \mathcal{P}(M)(X, Y)$  are the same as part a) and part b) in the proof of the above theorem.  $\{f_n\}$  is a contravariant Cauchy sequence, in order to show that  $\text{dist}(f_k, f) \rightarrow 0$ , as  $k \rightarrow \infty$ , just note, in part c) of the proof of above theorem, for any  $\epsilon > 0$ ,  $\exists N(\frac{\epsilon}{2})$ ,  $\text{dist}(f_k, f_p) < \frac{\epsilon}{2}$  only when  $p > k > N(\frac{\epsilon}{2})$ . Thus, for any  $x \in X, B \in \mathcal{B}$ ,

$$\begin{aligned} \frac{f_k(x,B)}{f_p(x,B)} &\geq \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f_k(x,B)}{f_p(x,B)} > e^{-\frac{\epsilon}{2}}, \text{ and } \lim_{p \rightarrow \infty} \frac{f_k(x,B)}{f_p(x,B)} = \frac{f_k(x,B)}{\liminf_{p \rightarrow \infty} f_p(x,B)} \\ &= \frac{f_k(x,B)}{f(x,B)} \geq e^{-\frac{\epsilon}{2}}. \text{ This gives us, } \text{dist}(f_k, f) \leq \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

If  $\{f_n\}$  is a covariant Cauchy sequence, then for  $\epsilon > 0$ ,  $\exists N(\frac{\epsilon}{2})$ ,  $\text{dist}(f_p, f_k) < \frac{\epsilon}{2}$ , when  $p > k > N(\frac{\epsilon}{2})$ . Thus, for any  $x \in X$ ,  $B \in \mathcal{B}$ ,

$$\begin{aligned} \frac{f_p(x,B)}{f_k(x,B)} &\geq \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{f_p(x,B)}{f_k(x,B)} > e^{-\frac{\epsilon}{2}}, \text{ and } \lim_{p \rightarrow \infty} \frac{f_p(x,B)}{f_k(x,B)} = \frac{\liminf_{p \rightarrow \infty} f_p(x,B)}{f_k(x,B)} \\ &= \frac{f(x,B)}{f_k(x,B)} \geq e^{-\frac{\epsilon}{2}}. \text{ This gives us, } \text{dist}(f, f_k) \leq \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

Finally, we discuss compactness.

We already know that  $\mathcal{P}(m)(1, n) = \mathcal{P}(n)$  is compact (contravariant topology).

proposition 5.  $(Y, \mathcal{B})$  is a measurable space. If  $\mathcal{P}(m)(1, Y) = \mathcal{P}(Y)$  with contravariant topology is compact, then  $\mathcal{P}(m)(n, Y)$  with contravariant topology is compact, where  $n$  is a  $n$ -elements discrete space.

Proof. Since  $n$  is discrete, so  $\mathcal{P}(m)(n, Y) \approx (\mathcal{P}(m)(1, Y))^n = \mathcal{P}(Y)^n$  as sets. With product topology on  $\mathcal{P}(m)(n, Y) \approx \mathcal{P}(Y)^n$  (each component  $\mathcal{P}(Y)$  with contravariant topology),  $\mathcal{P}(m)(n, Y)$  is compact.

Let  $B(\epsilon, f)$  be an open ball in  $\mathcal{P}(m)(n, Y)$ . If  $g \in B(\epsilon, f)$ , then  $g(i, -)$ ,  $i \in n$ , is a probability measure of  $(Y, \mathcal{B})$  and

$\text{dist}(g, f) = -\log \inf_{\substack{i \in \mathbb{N} \\ B \in \mathcal{B}}} \frac{g(i, B)}{f(i, B)} < \varepsilon$ . So for each  $i \in \mathbb{N}$ ,

$$\inf_{B \in \mathcal{B}} \frac{g(i, B)}{f(i, B)} \geq \inf_{\substack{i \in \mathbb{N} \\ B \in \mathcal{B}}} \frac{g(i, B)}{f(i, B)} > e^{-\varepsilon}, \text{ i.e. } \text{dist}(g(i, -), f(i, -)) < \varepsilon.$$

Thus,  $g \in \prod_{i=1}^{\infty} B(\varepsilon, f(i, -))$ . Conversely,  $(g_i) \in \mathcal{P}(Y)^{\mathbb{N}}$ , if  $(g_i) \in \prod_{i=1}^{\infty} B(\varepsilon, f(i, -))$ , denote  $g_i$  by  $g(i, -)$  and  $(g_i)$  by  $g$ . Since  $\inf_{B \in \mathcal{B}} \frac{g(i, B)}{f(i, B)} > e^{-\varepsilon}$  for

all  $i \in \mathbb{N}$ , so  $\text{dist}(g, f) = -\log \inf_{\substack{i \in \mathbb{N} \\ B \in \mathcal{B}}} \frac{g(i, B)}{f(i, B)} < \varepsilon$ . Thus, we have

$$B(\varepsilon, f) = \prod_{i=1}^{\infty} B(\varepsilon, f(i, -)).$$

But  $\prod_{i=1}^{\infty} B(\varepsilon, f(i, -))$  is an open set in the product topology on  $\mathcal{P}(m)(n, Y) \approx \mathcal{P}(Y)^{\mathbb{N}}$ , therefore the contravariant topology is contained in the product topology on  $\mathcal{P}(m)(n, Y)$ , and the contravariant topology on  $\mathcal{P}(m)(n, Y)$  is compact.

Corollary. The contravariant topology on  $\mathcal{P}(m)(n, m)$  is compact.

Remark. Generally,  $\mathcal{P}(m)(X, Y)$  is just a subspace of  $\mathcal{P}(Y)^X$ . The open ball  $B(\varepsilon, f)$  is too small, like open sets in box topology:

$$\text{because } g \in B(\varepsilon, f), \sup_{x \in X} \text{dist}(g(x, -), f(x, -)) = \text{dist}(g, f) < \varepsilon,$$

i.e. on each  $x$ -component,  $B(\varepsilon, f)$  is bounded by  $\varepsilon$ . If  $X$  is not finite, we can not find an open set (in product topology of  $\mathcal{P}(Y)^X$ ) contained in  $B(\varepsilon, f)$ .

In the case of usual metric space, completeness and total boundedness implies compactness, but since the generalized metric

is not symmetric, the usual proof does not work here. We have to put some new conditions together with completeness and total boundedness to imply countable compactness.

$A$  is a generalized metric space.

Definition 5.  $A$  is totally bounded(contravariant), if for any  $\varepsilon > 0$ , there exist finitely many  $B(\varepsilon, a_i)$ ,  $a_i \in A$ , covering  $A$ . Similarly we can define totally bounded(covariant).

Definition 6. Suppose that for any sequence  $\{a_i\}$  in  $A$ , if for some  $a \in A$ ,  $B(\varepsilon, a) \cap \{a_i\}$  is infinite, then there exists  $a_{i_0} \in \{a_i\}$  such that  $B(\varepsilon, a_{i_0}) \cap \{a_i\}$  is also infinite. We say that  $A$  has the property I (contravariant).

Similarly, we can define the property I (covariant).

Remark. In the usual case, total boundedness implies property I.

Lemma. If  $A$  is a generalized metric space, then, in the contravariant topology:

- a).  $A$  is first countable,
- b). sequential compactness is equivalent to countable compactness,
- c). compactness implies countable compactness, but the converse may not be true.

Proof. a). any  $a \in A$ ,  $\{B(r, a) \mid r \text{ rational}\}$  form a countable basis of neighbourhoods at  $a$ .

b). Roydon, Real Analysis, p.160.

c). the reason of the possible failure is:

even both  $a, b \in B(\varepsilon, c)$ , i.e.  $\text{dist}(a, c) < \varepsilon$ ,  $\text{dist}(b, c) < \varepsilon$ , but

$\text{dist}(a,b)$  still could be arbitrary, as in the following example,

$$A = \mathcal{P}(m)(1,3), \quad c = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad a = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix}, \quad b = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{bmatrix}$$

$a, b \in B(1,c)$ , since  $\text{dist}(a,c) = -\log \frac{1}{2} = \log 2 < 1$ ,  $\text{dist}(b,c) = -\log \frac{1}{2} = \log 2 < 1$ , but  $\text{dist}(a,b) = \infty$ .

Proposition 6. If  $A$  is complete(contravariant), totally bounded(covariant) and has property I(covariant), then  $A$  is countably compact(contravariant topology).

Proof. We are going to show that  $A$  is sequential compact (contravariant), then we could use the Lemma.

Let  $\{a_j\}$  be a sequence in  $A$ . For  $\varepsilon_1 = 1$ ,  $\exists B(b_{1_i}, 1)$ ,  $i=1, \dots, n_1$ ,

cover  $A$ , since  $A$  is totally bounded(covariant). Among them, there is  $B(b_{1_{i_0}}, 1)$ , such that  $B(b_{1_{i_0}}, 1) \cap \{a_j\}$  is infinite.

Using the property I, there exists  $a_{j_1} \in \{a_j\}$ , and  $B(a_{j_1}, 1) \cap \{a_j\}$

is infinite. For  $\varepsilon_2 = \frac{1}{2}$ ,  $\exists B(b_{2_i}, \frac{1}{2})$ ,  $i=1, \dots, n_2$ , cover  $A$ . Then

$\exists B(b_{2_{i_0}}, \frac{1}{2})$ , such that  $B(b_{2_{i_0}}, \frac{1}{2}) \cap (B(a_{j_1}, 1) \cap \{a_j\})$  is infinite,

then by property I,  $\exists a_{j_2} \in B(a_{j_1}, 1) \cap \{a_j\}$ , such that

$B(a_{j_2}, \frac{1}{2}) \cap (B(a_{j_1}, 1) \cap \{a_j\})$  is infinite. .... For  $\varepsilon_m = \frac{1}{m}$ ,

$\exists B(b_{m_i}, \frac{1}{m})$ ,  $i=1, \dots, n_m$ , cover  $A$ . Then  $\exists B(b_{m_{i_0}}, \frac{1}{m})$ ,

$B(b_{m_{i_0}}, \frac{1}{m}) \cap (\bigcap_{k=1}^{m-1} B(a_{j_k}, \frac{1}{k}) \cap \{a_j\})$  is infinite, by property I,

$$\exists a_{j_m} \in \bigcap_{k=1}^{m-1} B(a_{j_k}, \frac{1}{k}) \cap \{a_j\} \quad \text{and} \quad B(a_{j_m}, \frac{1}{m}) \cap (\bigcap_{k=1}^{m-1} B(a_{j_k}, \frac{1}{k}) \cap \{a_j\})$$

is infinite. Continuing this process, we get  $\{a_{j_m}\}$ , a sub-

sequence of  $\{a_j\}$ , and  $\{a_{j_m}\}$  is a contravariant Cauchy sequence:

for any  $\varepsilon > 0$ ,  $\exists N > 0$ , such that  $\frac{1}{N} < \varepsilon$ , then for  $p > k > N$ ,

$$a_{j_p} \in B(a_{j_k}, \frac{1}{k}), \text{ thus, } \text{dist}(a_{j_k}, a_{j_p}) < \frac{1}{k} < \frac{1}{N} < \varepsilon.$$

Since  $A$  is complete(contravariant), thus,  $\{a_{j_m}\}$  has a limit.

Remark. If we interchange the contravariant and covariant in the above proposition, we will get conditions for  $A$  being countably compact (covariant topology).

CHAPTER 3. APPLICATIONS IN STATISTICS,  
STOCHASTIC PROCESSES AND STOCHASTIC DYNAMIC PROGRAMMING

There are four sections in this chapter which deal with the formulation of statistical decision problems and the existence of the best decision rules; the formulation of stochastic processes and transformations, stochastic and Markov dynamics, and the relations between them; decomposition of stochastic transformations; formulation of stochastic dynamic programming and the existence of the optimal policies.

§1 Statistical decision problems.

1). The usual structure of statistical decision problems may be described by the following structure:

$(X, \mathcal{A})$	the message space
$(\Omega, \mathcal{G})$	the space of observable states
$(Y, \mathcal{B})$	the decision space

We may use a random map  $f: X \rightarrow \Omega$  to describe a noisy transmission channel: for all  $x \in X$ ,  $C \in \mathcal{G}$ ,  $f(x, C)$  gives the probability that we observe states lying in  $C$  when message  $x$  is sent.

A procedure for making decision depending on what has been observed is called a decision rule,  $\delta: \Omega \rightarrow Y$ . It also may be random (probabilistic).

A criterion is needed to measure the decision rules. Sometimes a direct relation  $g: X \rightarrow Y$  between  $X$  and  $Y$  has been assumed, (see James L. Melsa and David L. Cohn, Decision and Estimation Theory). But in the most of the cases, a two variable real-valued

cost function  $c: X \times Y \rightarrow R$  has been used.  $c(x,y)$  gives the cost for us to choose decision  $y$  if message  $x$  be sent.

The problem is , given  $f: X \rightarrow \Omega$  and some connection between  $X$  and  $Y$ , how could we find a "best" decision rule (optimal decision rule)?

a). The relation between  $X$  and  $Y$ .

The usual relation between  $X$  and  $Y$  is given by a cost function. There is no general method to give a cost function, like the often-used "mean square cost function" in statistics books. People just use it without giving reason for choosing this rather than some other functions.

Cost function is a measure of the difference between certain interesting characteristics of the real process and the decision we make about these characteristics according to the observations (based on what we think is happening). It is quite often that cost functions have been constructed just "for analytical purposes". (see Cencov).

Here, instead of using cost function, we consider a "correct action" (as in Melsa and Cohn). This is the direct connection between  $X$  and  $Y$ . The correct choice for the decisions exists objectively in the process itself, we may get an approximation of it through experiments, analysing, synthesizing the observed data, and certain consideration to simplify the calculation. The correct action  $g: X \rightarrow Y$  may be random too.

b). The criterion of "best".

$\mathcal{P}(M)(\Omega, Y)$  is the space of all decision rules.

If we have a cost function  $c: X \times Y \rightarrow K$  (where  $K$  is the nonnegative reals), then define

$$c_{\delta}: \mathcal{P}(m)(\Omega, Y) \rightarrow K^X, \quad \delta \mapsto c_{\delta},$$

$$\text{where } c_{\delta}(x) = \int_Y c(x, y) \delta \circ f(x, dy), \quad x \in X;$$

$$\text{also define } \sup_x(\ ): K^X \rightarrow K, \quad g \mapsto \sup_x g(x).$$

We have the following diagram

$$\begin{array}{ccc} \mathcal{P}(m)(\Omega, Y) & \xrightarrow{c_{\delta}} & K^X \\ & \searrow & \downarrow \sup_x(\ ) \\ & \sup_x c_{\delta} & K \end{array}$$

the composition sends  $\delta$  in  $\mathcal{P}(m)(\Omega, Y)$  to  $\sup_x c_{\delta}(x)$ .

Certainly we want our decision rule  $\delta$  to make the loss under the worst case,  $\sup_x c_{\delta}(x)$ , as small as possible.

If we have a priori probability measure  $p$  on the message space  $X$ ,  $p: 1 \mapsto X$ , then instead of using  $\sup_x(\ )$ , just take the

functional average with respect to  $p$ ,  $\int_X g dp$ , where  $g \in K^X$ .

$$\begin{array}{ccc} \mathcal{P}(m)(\Omega, Y) & \xrightarrow{c_{\delta}} & K^X \\ & \searrow & \downarrow \int_X (\ ) dp \\ & & K \end{array}$$

$$\delta \mapsto c_{\delta} = \int_Y c(x, y) \delta \circ f(x, dy) \mapsto \int_X \int_Y c(x, y) \delta \circ f(x, dy) dp = c_{\delta}^p,$$

Similarly, we want  $c_0^D$  as small as possible.

If we use reward function  $r: X \times Y \rightarrow K$ ,  $r(x,y)$  gives the reward for choosing decision  $y$  when message  $x$  is sent. Then, we would use  $\inf_x$ , so that in the worst case, we still could earn as much as possible.

Example.  $X, Y$  and  $\Omega$  are finite sets, let  $A = \mathcal{P}(X)$ ,  $B = \mathcal{P}(Y)$  and  $C = \mathcal{P}(\Omega)$ . For any  $f: X \rightarrow Y$  in  $\mathcal{P}(m)$ ,  $f_{*-}: A = \mathcal{P}(X) \rightarrow B = \mathcal{P}(Y)$  is a morphism in  $\mathcal{C}$ . Conversely, given any  $g: A \rightarrow B$  define  $\bar{g}(x, -) = g(\delta_x)$ , where  $\delta_x$  is the probability measure which is concentrated on  $x \in X$ . Thus, we have the bijection

$$\frac{X \rightarrow Y \text{ in } \mathcal{P}(m)}{A \rightarrow B \text{ in } \mathcal{C}}$$

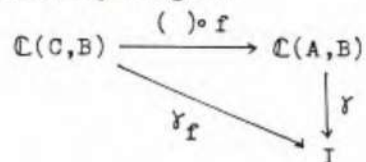
or  $\mathcal{P}(m)(X, Y) \approx \mathcal{C}(A, B)$  as sets. Similarly we can get the space of decision rules  $\mathcal{P}(m)(\Omega, Y) \approx \mathcal{C}(C, B)$ .

If  $c: X \times Y \rightarrow K$  is a cost function,  $p$  is a priori probability measure on  $X$ . Define

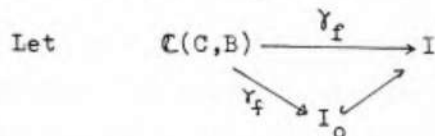
$$\begin{aligned} \gamma: \mathcal{P}(m)(X, Y) \approx \mathcal{C}(A, B) &\longrightarrow K \\ g &\longmapsto \int \left( \int c(x, y) g(x, dy) \right) dp \\ &= \sum_{\substack{x \in X \\ y \in Y}} c(x, y) g(x, y) p(x) && \text{(since } X, Y \text{ are finite discrete)} \\ &\leq \sum_{\substack{x \in X \\ y \in Y}} c(x, y) && (0 \leq g(x, y), p(x) \leq 1) \end{aligned}$$

then  $\gamma$  is affine linear and  $\gamma$  is bounded, that is  $\gamma: \mathcal{C}(A, B) \rightarrow I$  for some closed interval  $I$  in  $K$ .

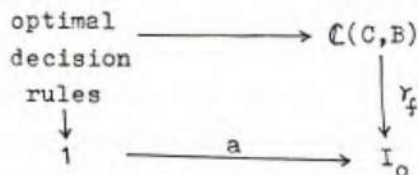
In the following diagram



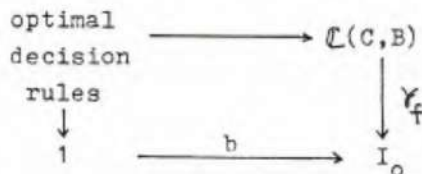
$$\gamma_f = \gamma \circ ((\ ) \circ f), \quad \gamma_f(s) = \gamma(s \circ f) = c_s^p.$$



be the image factorization of  $\gamma_f$  in  $\mathcal{C}$ . Assume  $\inf I_0 = a$ ,  $\sup I_0 = b$ , then the pullback along  $1 \xrightarrow{a} I_0$  gives the optimal decision rules when we take  $c$  as the cost function and  $p$  as a priori probability measure.

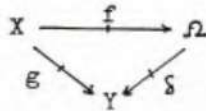


If  $c$  is the reward function, the pullback along  $1 \xrightarrow{b} I_0$  gives the optimal decision rules



Secondly, the metric criterion.

Since  $\mathcal{P}(m)$  is enriched in  $\mathcal{U}\text{-cat}$  (generalized metric spaces), we may use this generalized metric  $\text{dist}$  as a measure of "best". In the following diagram,



$g$  is the correct action,  $\delta$  is a decision rule. Then, we want the distance between the composition  $\delta \circ f$  and  $g$  as small as possible. If  $\delta^*: \Omega \rightarrow Y$  satisfies

$$\text{dist}(\delta^* \circ f, g) \leq \text{dist}(\delta \circ f, g) \quad \text{for all } \delta \in \mathcal{P}(m)(\Omega, Y)$$

$$\text{or} \quad \text{dist}(\delta^* \circ f, g) = \inf_{\delta} \text{dist}(\delta \circ f, g)$$

then we say that  $\delta^*$  is a best decision rule.

2). Comparison of the cost function criterion and metric criterion.

Let  $X$  be a measurable space,  $|(X|\mathcal{P}(m))|$  is the set of objects of comma category  $(X|\mathcal{P}(m))$ , viewing it as a discrete category.

For any  $c: X \times Y \rightarrow K$ , assume  $c \in \mathcal{M}(X \times Y, K)$ , and  $f: X \rightarrow \Omega$ ,  $f \in |(X|\mathcal{P}(m))|$ , define

$$c_f = \left\{ \delta_0 \in \mathcal{P}(m)(\Omega, Y) \mid \inf_{\delta} \sup_x \int_Y c(x, y) \delta \circ f(x, dy) = \sup_x \int_Y c(x, y) \delta_0 \circ f(x, dy) \right\}$$

$$c^f = \left\{ \delta_0 \in \mathcal{P}(m)(\Omega, Y) \mid \sup_{\delta} \inf_x \int_Y c(x, y) \delta \circ f(x, dy) = \inf_x \int_Y c(x, y) \delta_0 \circ f(x, dy) \right\}$$

$c_f$  is the set of optimal solutions with respect to the cost function  $c$ ;  $c^f$  is the set of optimal solutions with respect to the reward function  $c$ .

For each  $f: X \rightarrow \Omega$ ,  $g: X \rightarrow Y$ , define

$$g_f = \{ \delta_0 \in \mathcal{P}(m)(\Omega, Y) \mid \text{dist}(\delta_0 \circ f, g) = \inf_{\delta} \text{dist}(\delta \circ f, g) \}$$

$g_f$  is the set of optimal solutions with respect to the "correct action"  $g$ .

$c_f$ ,  $c^f$ , and  $g_f$  are functors from  $|X| \mathcal{P}(m)|$  to Set.

Question (1). For given  $c$ , is there some  $g: X \rightarrow Y$ , such that  $c_f = g_f$  or  $c^f = g_f$  for all  $f \in |X| \mathcal{P}(m)|$  ;

Question (2). For a given  $g$ , is there some  $c: X \times Y \rightarrow K$ , such that  $g_f = c_f$  or  $g_f = c^f$  for all  $f \in |X| \mathcal{P}(m)|$  ,

or we could ask inclusion instead of equality, eg.  $c_f \subseteq g_f$ , etc.

Proposition 1. If  $g$  is deterministic (i.e.  $g$  is a measurable function), then there is a  $c: X \times Y \rightarrow K$ , such that  $g_f = c^f$ , for all  $f \in |X| \mathcal{P}(m)|$  .

Proof. Define  $c: X \times Y \rightarrow K$ , as  $c(x, y) = \begin{cases} 1, & \text{if } y = g(x) \\ 0, & \text{if } y \neq g(x) \end{cases}$ .

Then,  $\int_Y c(x, y) \delta \circ f(x, dy) = \delta \circ f(x, g(x))$ ,  $f \in \mathcal{P}(m)(X, \Omega)$ ,  $\delta \in \mathcal{P}(m)(\Omega, Y)$

hence,  $\inf_x \int_Y c(x, y) \delta \circ f(x, dy) = \inf_x \delta \circ f(x, g(x))$ ;

and  $\text{dist}(\delta \circ f, g) = -\log \inf_{\substack{x \in X \\ B \in \mathcal{B}}} \frac{\delta \circ f(x, B)}{g(x, B)} = -\log \inf_x \delta \circ f(x, g(x))$ .

For any  $\delta_0 \in \mathcal{P}(m)(\Omega, Y)$ ,

$$\delta_0 \in c^f \text{ iff } \sup_{\delta} \inf_x \delta \circ f(x, g(x)) = \inf_x \delta_0 \circ f(x, g(x)) \quad (1)$$

$$\text{iff } -\log \sup_{\delta} \inf_x \delta \circ f(x, g(x)) = -\log \inf_x \delta_0 \circ f(x, g(x)) \quad (2)$$

$$\text{iff } \inf_{\delta} \text{dist}(\delta \circ f, g) = \text{dist}(\delta_0 \circ f, g), \quad (3)$$

$$\text{iff } \delta_0 \in \mathcal{G}_f. \quad (4)$$

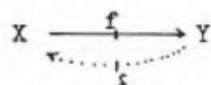
(the left side of (2)  $-\log \sup_{\delta} \inf_x \delta \circ f(x, g(x)) = \inf_{\delta} (-\log \inf_x \delta \circ f(x, g(x)))$ )

$= \inf_{\delta} \text{dist}(\delta \circ f, g)$ ; and the right side of (2)

$$-\log \inf_x \delta_0 \circ f(x, g(x)) = \text{dist}(\delta_0 \circ f, g)$$

3). Parameter estimate.

In this case,  $X$  is the space of parameters and  $Y = X$ ,  $g = 1_X$ ,



then we want  $\text{dist}(\delta \circ f, 1_X)$  as small as possible, under some conditions,  $\delta \circ f$  may even be equal to  $1_X$ .

Remark. The identity  $1_X$  in  $\mathcal{P}(\mathcal{M})$  is a random map defined as

$$1_X(x, A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}. \text{ Note that, the identity } 1_X \text{ in } \mathcal{M} \text{ gives rise}$$

to  $1_X$  in  $\mathcal{P}(\mathcal{M})$ , but the converse may not be true, i.e.  $1_X$  in  $\mathcal{P}(\mathcal{M})$

is not necessary coming from  $1_X$  in  $\mathcal{M}$  (see next remark).

Proposition 2.  $X, Y \in \mathcal{M}$ , with  $\sigma$ -algebra  $\mathcal{A}$  and  $\mathcal{B}$ ,  $f: X \rightarrow Y$  is deterministic, and a).  $f$  is injective, b).  $f(A) \in \mathcal{B}$ , for any

$A \in \mathcal{A}$ . Then there is a  $\delta: Y \rightarrow X$  in  $\mathcal{M}$ , such that

$$\delta \circ f = 1_X \text{ in } \mathcal{M}, \quad \text{so is } 1_X \text{ in } \mathcal{P}(\mathcal{M}).$$

Proof. Define  $\delta: Y \rightarrow X$ ,  $\delta(y) = \begin{cases} x, & \text{if } f(x) = y \\ x_0, & \text{if } y \notin f(X), x_0 \text{ is a fixed} \\ & \text{element in } X \end{cases}$

$\delta$  is well defined, since  $f$  is injective.  $\delta$  is measurable, since for each  $A \in \mathcal{A}$ ,

$$\delta^{-1}(A) = \begin{cases} f(A), & x_0 \notin A \\ f(A) \cup \{Y - f(A)\}, & x_0 \in A \end{cases}$$

and  $\delta \circ f(x) = x$ , hence  $\delta \circ f = 1_X$  in  $\mathcal{M}$ , so is in  $\mathcal{P}(\mathcal{M})$ .

Definition.  $X$  is a separable measurable space, if for any  $x \in X$ , there is a minimal measurable set  $E_x$  containing  $x$  (minimal in the sense that it contains no proper sub measurable set).

Proposition 3.  $X$  is a separable measurable space, define  $x \sim x'$  iff  $E_x = E_{x'}$ . Then  $X \approx X/\sim$  in  $\mathcal{P}(\mathcal{M})$ .

Proof. Let  $f: X \rightarrow X/\sim$  be the quotient map in  $\mathcal{M}$ , where  $X/\sim$  with  $\sigma$ -algebra  $\bar{\mathcal{A}}$ ,  $\bar{A} \in \bar{\mathcal{A}}$  iff  $f^{-1}(\bar{A}) \in \mathcal{A}$ . Define  $g: X/\sim \rightarrow X$  just as a section of  $f$ , i.e.  $f \circ g = 1_{X/\sim}$  in Set.

For any  $A \in \mathcal{A}$ ,  $g^{-1}(A)$  is measurable if  $f \circ g^{-1}(A) = (g \circ f)^{-1}(A) \in \mathcal{A}$ , and since  $g \circ f(x) \sim x$ , so  $E_x = E_{g \circ f(x)}$ . Thus,  $x \in A$  iff  $g \circ f(x) \in A$ , i.e.  $A = g \circ f(A)$ , or  $(g \circ f)^{-1}(A) = A$ . This shows that  $g$  is measurable. So  $f \circ g = 1_{X/\sim}$  is in  $\mathcal{M}$ , hence is in  $\mathcal{P}(\mathcal{M})$ .

$$g \circ f(x, A) = \begin{cases} 1, & \text{if } x \in A \text{ (since } g \circ f(x) \in A \text{ iff } x \in A) \\ 0, & \text{otherwise} \end{cases}$$

this is  $g \circ f = 1_X$  in  $\mathcal{P}(\mathcal{M})$ , therefore  $X \approx X/\sim$  in  $\mathcal{P}(\mathcal{M})$ .

Remark.  $X$  is a separable measurable space,  $f: X \rightarrow X$  is a

measurable function. If  $f(E_x) \subseteq E_x$  for all  $x \in X$ , then  $f = 1_X$  in  $\mathcal{P}(m)$ .

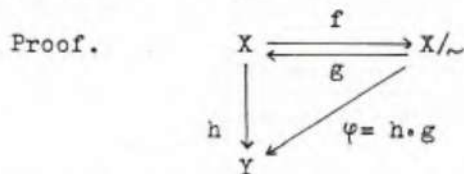
Proof.  $x \in E_x$  implies  $f(x) \in f(E_x)$ , from the assumption  $f(E_x) \subseteq E_x$ , so we have  $f(x) \in E_x$ . Thus,  $E_{f(x)} \subseteq E_x$ , but  $E_x$  is a minimal measurable set, so  $E_{f(x)} = E_x$ .

Now, for  $\forall A \in \mathcal{A}$ ,  $x \in A$  iff  $E_x \subseteq A$  iff  $E_{f(x)} \subseteq A$  iff  $f(x) \in A$ , so

$$f(x, A) = \begin{cases} 1, & f(x) \in A \\ 0, & f(x) \notin A \end{cases} = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

i.e.  $f = 1_X$  in  $\mathcal{P}(m)$ .

Proposition 4.  $X$  is a separable measurable space,  $h: X \rightarrow Y$  is a measurable function. If a).  $h^{-1}(h(E_x)) \subseteq E_x$ ; b).  $\exists$  a section  $g$  of  $h$ , such that  $h \circ g(\bar{A})$  is measurable for all  $\bar{A} \in \bar{\mathcal{A}}$  (where  $f: X \rightarrow X/\sim$ ), then there is a measurable function  $\delta: Y \rightarrow X$ , such that  $\delta \circ h = 1_X$  in  $\mathcal{P}(m)$ .



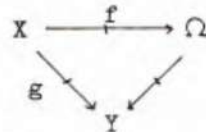
Let  $\varphi = h \circ g$ , then  $h = \varphi \circ f$  and  $\varphi$  is injective, since a). From proposition 2, there is  $\delta': Y \rightarrow X/\sim$ , such that  $\delta' \circ \varphi = 1_{X/\sim}$ . Define  $\delta = g \circ \delta'$ , then  $\delta \circ h = g \circ \delta' \circ \varphi \circ f = g \circ (\delta' \circ \varphi) \circ f = g \circ (1_{X/\sim}) \circ f = g \circ f = 1_X$ .

4). Optimal decision rules under the metric criterion.

Proposition 5.

$\text{dist}(\cdot, f, g): \mathcal{P}(m)(\Omega, Y) \longrightarrow [0, \infty]$

is a Lip 1 function, hence is a continuous function.



Proof. First,  $\cdot, f: \mathcal{P}(m)(\Omega, Y) \longrightarrow \mathcal{P}(m)(X, Y)$  is Lip 1:

Let  $h_1, h_2, h_3 \in \mathcal{P}(m)(\Omega, Y)$ , and  $h_1 = th_2 + (1-t)h_3$ , where  $t \in [0, 1]$ ,

then  $h_1 \circ f = th_2 \circ f + (1-t)h_3 \circ f$ ,

and  $\text{dist}(h_1, h_2) = \inf_{\substack{\exists t \in [0, 1], \exists h_3 \in \mathcal{P}(m)(\Omega, Y) \\ h_1 = th_2 + (1-t)h_3}} (-\log t)$

$$\geq \inf_{\substack{\exists s \in [0, 1], \exists h' \in \mathcal{P}(m)(X, Y) \\ h_1 \circ f = sh_2 \circ f + (1-s)h'}} (-\log s) = \text{dist}(h_1 \circ f, h_2 \circ f).$$

Secondly,  $g \in \mathcal{P}(m)(X, Y)$ , then  $\text{dist}(\cdot, g): \mathcal{P}(m)(X, Y) \longrightarrow [0, \infty]$

is a Lip 1 function:

Because,  $\mathcal{U} = ([0, \infty], \geq)$  is a closed category, so

$$\frac{\text{dist}(g_1, g_2) + \text{dist}(g_2, g) \geq \text{dist}(g_1, g)}{\text{dist}(g_1, g_2) \geq [\text{dist}(g_2, g), \text{dist}(g_1, g)]}$$

where  $g_1, g_2 \in \mathcal{P}(m)(X, Y)$ ,  $[-, -]$  is the Hom in  $\mathcal{U}$ . Note that,  $\text{dist}(\cdot, g)$  is a contravariant representable  $\mathcal{U}$ -functor from  $\mathcal{U}$ -category  $\mathcal{P}(m)(X, Y)$  to  $\mathcal{U}$ .

Now,  $\text{dist}(\cdot, f, g)$  is the composition of two Lip 1 functions, so it is a Lip 1 function, hence a continuous function.

Corollary. If  $\mathcal{P}(m)(\Omega, Y)$  is countably compact, then there is a best decision rule  $\delta^*$ , such that  $\text{dist}(\delta^* \cdot, f, g) = \inf_{\delta} \text{dist}(\delta \cdot, f, g)$ .

§2. Stochastic processes, Stochastic transformations and  
Stochastic dynamics

A stochastic process is a sequence of probability distributions on the accumulated state spaces,  $p_n: 1 \rightarrow \prod_{k=1}^n X_k$ ,  $n = 1, 2, \dots$ , where  $X_k$  is the state space at stage  $k$ ,  $p_n$ 's are compatible, i.e. the following diagram is commutative for each  $n$ ,

$$\begin{array}{ccc}
 1 & \xrightarrow{p_{n+1}} & \prod_{k=1}^{n+1} X_k \\
 & \searrow p_n & \downarrow \pi_{(n)} \\
 & & \prod_{k=1}^n X_k
 \end{array}$$

where  $\pi_{(n)}$  is the projection.

Remark. Most often, the definition of a stochastic process is given as a sequence of real valued measurable functions (random variables)  $\{f_n\}_{n=1,2,\dots}$ , then the joint distribution  $F_n$  of  $f_1, f_2, \dots, f_n$  gives a probability distribution on  $R^n$ , where  $R$  is the reals, the diagram

$$\begin{array}{ccc}
 1 & \xrightarrow{F_{n+1}} & R^{n+1} \\
 & \searrow F_n & \downarrow \pi_{(n)} \\
 & & R^n
 \end{array}$$

is commutative for each  $n$ .

This definition could be generalized as follows:

Definition 1. A stochastic process consists of state spaces  $X_n$ ,  $n=1, 2, \dots$ ,  $X_n$  contains accumulated states until stage  $n$ ;

recall of history  $h_n: X_{n+1} \rightarrow X_n$ ,  $h_n$  is deterministic, i.e. it is a measurable function; and a sequence of probability measures  $P_n: 1 \rightarrow X_n$ , such that for each  $n$ , the diagram

$$\begin{array}{ccc}
 1 & \xrightarrow{P_{n+1}} & X_{n+1} \\
 & \searrow P_n & \downarrow h_n \\
 & & X_n
 \end{array}$$

is commutative.

Example. Taking  $X_n$  as a subspace of the product space in the previous case, such that the restricted projection  $\pi_{(n)}$  are compatible with  $X_n$ 's.

More generally, we have

Definition 2. The category of stochastic transformations  $\mathcal{P}(\mathcal{M}^{\text{st}})$ : object  $X$  is a sequence of measurable spaces  $X_n$ ,  $n=1,2,\dots$ , together with recall of history  $h_n: X_{n+1} \rightarrow X_n$  in  $\mathcal{M}$ , we denote

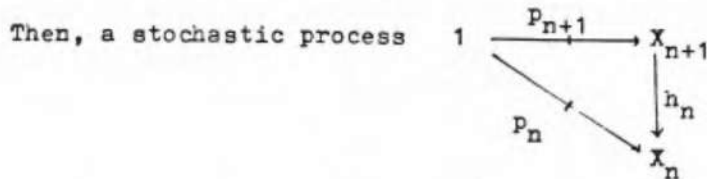
$$(X_1 \xleftarrow{h_1} X_2 \leftarrow \dots \leftarrow X_n \xleftarrow{h_n} X_{n+1} \leftarrow \dots)$$
 by  $(X, h)$ ;

morphism  $\varphi: (X, h) \rightarrow (Y, g)$  is just a natural transformation from  $(X, h)$  to  $(Y, g)$ , i.e. for each  $n$ ,  $\varphi_n: X_n \rightarrow Y_n$  is in  $\mathcal{P}(\mathcal{M})$ ,

and the square

$$\begin{array}{ccc}
 X_{n+1} & \xrightarrow{\varphi_{n+1}} & Y_{n+1} \\
 \downarrow h_n & & \downarrow g_n \\
 X_n & \xrightarrow{\varphi_n} & Y_n
 \end{array}$$

is commutative.



$n = 1, 2, \dots$ , is just a stochastic transformation  $p$  from one point space  $1$  to  $(X, h)$ ; and any  $\varphi: (X, h) \rightarrow (Y, g)$  will transform a stochastic process  $p$  on  $(X, h)$  to a stochastic process  $\varphi \cdot p$  on  $(Y, g)$ ,  $(\varphi \cdot p)_n = \varphi_n \cdot p_n: 1 \rightarrow Y_n$ .

Remark. a). The above definition of stochastic process places emphasis on the results of dynamics (motions). If we take  $p_n: 1 \rightarrow X_n$  as a point in  $\mathcal{P}(X_n)$  (the probability measures on  $X_n$ ) then a stochastic process  $p$  is just a path in space  $(X, h)$ .

b). The set  $\mathcal{P}(m^{wp})(X, Y)$  of stochastic transformations between  $(X, h)$  and  $(Y, g)$  is a convex set, so it is a generalized metric space. If  $\varphi, \psi \in \mathcal{P}(m^{wp})(X, Y)$ , then

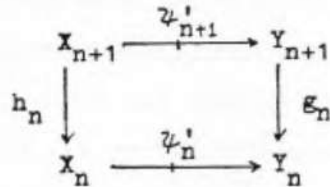
$$\text{dist}(\varphi, \psi) =$$

$$\begin{aligned}
 \inf_{\substack{t \in [0, 1], \exists \varphi'_n \in \mathcal{P}(m^{wp})(X, Y) \\ \varphi = t\varphi + (1-t)\varphi'}} (-\log t) & \geq \inf_{\substack{t_n \in [0, 1], \exists \varphi'_n \in \mathcal{P}(m)(X_n, Y_n) \\ \varphi_n = t_n \varphi_n + (1-t_n) \varphi'_n}} (-\log t_n) \\
 & = \text{dist}(\varphi_n, \varphi'_n)
 \end{aligned}$$

thus,  $\text{dist}(\varphi, \psi) \geq \sup_n \text{dist}(\varphi_n, \psi_n)$ .

Note that, a sequence of random maps  $\psi_n' \in \mathcal{P}(M)(X_n, Y_n)$ ,  $n = 1, 2, \dots$

is not necessarily compatible with recall of history  $h$  and  $g$ , i.e. the square



may not be commutative, so  $\{\psi_n'\}_{n=1,2,\dots}$  is not necessarily coming from a stochastic transformation.

There is another way to assign a distance between two stochastic transformations  $(\varphi, \psi \in \mathcal{P}(M)^{\text{stoch}}(X, Y))$ ,

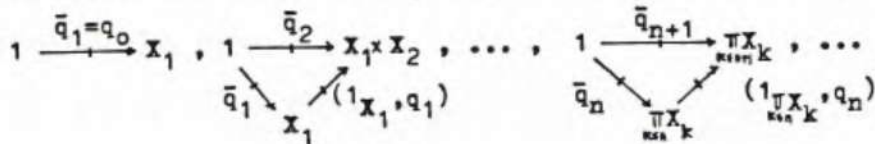
$$\overline{\text{dist}(\varphi, \psi)} = \sum_{n=1}^{\infty} \frac{1}{2^n} \text{dist}(\varphi_n, \psi_n), \text{ this is a generalized metric,}$$

since  $\text{dist}$  is.

Another way to analyse a stochastic process is to emphasize the dynamics itself, rather than its results. For the distinction, we call it stochastic dynamics.

Definition 3. A stochastic dynamics consists of a sequence of state spaces  $X_n$ ,  $n = 0, 1, \dots$ , where  $X_0 = 1$ ; and a sequence of dynamics (law of motion)  $q_n: \prod_{k=0}^{n-1} X_k \rightarrow X_n$ .

We can transform dynamics  $\{q_n\}$  into their results, a sequence of probability measures on each product space, as follows:



$\bar{q}_n$ ,  $n = 1, 2, \dots$  are compatible with projections  $\pi_{(n)} : \prod_{k=1}^n X_k \rightarrow \prod_{k=1}^{n-1} X_k$ .

Note that, if we only know the results of a sequence of dynamics, i.e. a sequence of probability measures on product spaces, generally we could not recover the dynamics, even when we can reconstruct it as in the following proposition, the choice of the dynamics is not unique.

Proposition 1.  $X, Y$  are finite sets,  $p: 1 \rightarrow X, q: 1 \rightarrow Y$  are two probability measures and  $f: X \rightarrow Y$  is a surjective deterministic map with the property  $f \cdot p = q$ . Then there is a random map  $\varphi: Y \rightarrow X$ , such that  $\varphi \cdot q = p$  and  $f \cdot \varphi = 1_Y$ .

Proof. Define  $\varphi: Y \rightarrow X$ ,

if  $q(y) \neq 0$ ,

$$\begin{aligned} \varphi(y, x) &= \frac{p(\{x\} \cap f^{-1}(y))}{p(f^{-1}(y))} = \frac{p(\{x\} \cap f^{-1}(y))}{q(y)} \\ &= \begin{cases} \frac{p(x)}{q(y)}, & \text{if } x \in f^{-1}(y) \\ 0, & \text{if } x \notin f^{-1}(y) \end{cases} ; \end{aligned}$$

if  $q(y) = 0$ , pick any  $x_y \in f^{-1}(y)$ ,

$$\varphi(y, x) = \begin{cases} 1, & x = x_y \\ 0, & x \neq x_y \end{cases} .$$

We have to check,

a).  $\varphi \in \mathcal{P}(m)(Y, X)$ :

for  $q(y) = 0$ ,  $\varphi(y, -)$  is a probability measure on  $X$  which is concentrated on  $x_y$  ;

for  $q(y) \neq 0$ ,  $\sum_{x \in X} \varphi(y, x) = \sum_{x \in X} \frac{p(\{x\} \cap f^{-1}(y))}{q(y)} = \frac{1}{q(y)} p(\cup_{x \in X} \{x\} \cap f^{-1}(y))$

$$= \frac{1}{q(y)} p(f^{-1}(y)) = \frac{q(y)}{q(y)} = 1,$$

thus,  $\varphi \in \mathcal{P}(m)(Y, X)$ .

b) .  $\varphi \cdot q = p$ :

let  $B_0 = \{y \mid q(y) = 0\}$ ,

then  $\varphi \cdot q(x) = \sum_{y \in Y} \varphi(y, x) q(y) = \sum_{B_0} \varphi(y, x) q(y) + \sum_{Y \setminus B_0} \varphi(y, x) q(y)$

$$= \sum_{Y \setminus B_0} \varphi(y, x) q(y) = \sum_{Y \setminus B_0} \frac{p(\{x\} \cap f^{-1}(y))}{q(y)} q(y)$$

$$= \sum_{Y \setminus B_0} p(\{x\} \cap f^{-1}(y)) = p(\{x\} \cap \bigcup_{Y \setminus B_0} f^{-1}(y)) = p(x)$$

the last equality holds, because  $q(y) = f \cdot p(y) = \sum_{x \in X} f(x, y) p(x) = \sum_{x \in f^{-1}(y)} p(x)$ , hence  $q(y) = 0$  implies  $p(x) = 0$  for any  $x \in f^{-1}(y)$ .

Since  $f$  is surjective,  $x \in f^{-1}(y)$ , some  $y \in Y$ . If  $y \in B_0$ , then  $p(x) = 0$ ,

and  $p(\{x\} \cap \bigcup_{Y \setminus B_0} f^{-1}(y)) \leq p(x) = 0$ , so  $p(\{x\} \cap \bigcup_{Y \setminus B_0} f^{-1}(y)) = 0 = p(x)$ ;

if  $y \notin B_0$ , then  $\{x\} \cap \bigcup_{Y \setminus B_0} f^{-1}(y) = \{x\}$ , and  $p(\{x\} \cap \bigcup_{Y \setminus B_0} f^{-1}(y)) = p(x)$ .

c) .  $f \cdot \varphi = 1_Y$ :

$$f \cdot \varphi(y, y') = \sum_{x \in X} f(x, y') \varphi(y, x) = \varphi(y, f^{-1}(y')) = \sum_{x \in f^{-1}(y')} \varphi(y, x)$$

$$\text{if } q(y) = 0, \quad \sum_{x \in f^{-1}(y')} \varphi(y, x) = \begin{cases} 1, & y = y' \\ 0, & y \neq y' \end{cases};$$

$$\begin{aligned} \text{if } q(y) \neq 0, \quad \sum_{x \in f^{-1}(y')} \varphi(y, x) &= \sum_{x \in f^{-1}(y')} \frac{p(\{x\} \cap f^{-1}(y))}{q(y)} = \frac{1}{q(y)} p(f^{-1}(y') \cap f^{-1}(y)) \\ &= \begin{cases} 1, & y = y' \\ 0, & y \neq y' \end{cases} \end{aligned}$$

therefore  $f \cdot \varphi = 1_Y$ .

Definition 4. If  $(X, q)$ ,  $(Y, p)$  are two stochastic dynamics as

in definition 3, a morphism  $f: (X,q) \longrightarrow (Y,p)$  consists of a sequence of random maps  $f_k: X_k \longrightarrow Y_k$ , such that the square

$$\begin{array}{ccc} \prod_{k \in \mathbb{N}} X_k & \xrightarrow{q_n} & X_{n+1} \\ \downarrow \otimes_{k \in \mathbb{N}} f_k & & \downarrow f_{n+1} \\ \prod_{k \in \mathbb{N}} Y_k & \xrightarrow{p_n} & Y_{n+1} \end{array}$$

is commutative for each  $n$ . If  $g: (Y,p) \longrightarrow (Z,p')$  is another morphism, since the tensor in  $\mathcal{P}(m)$  is a bifunctor, so  $g \cdot f$  with  $(g \cdot f)_n = g_n \cdot f_n$  is a morphism from  $(X,q)$  to  $(Z,p')$ . We denote this category as  $S_1$ .

Definition 5. A Markov dynamics (process) is a stochastic dynamics in which what will happen does not depend on the history, i.e.  $q_n: X_n \longrightarrow X_{n+1}$ ,  $n = 1, 2, \dots$ , where  $X_0 = 1$ . We define

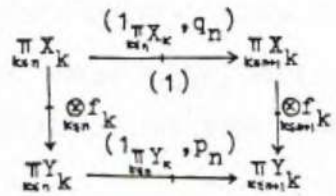
a morphism between two Markov dynamics similarly,  $g: (X,q) \longrightarrow (Y,p)$  consists of a sequence of random maps  $g_n: X_n \longrightarrow Y_n$ , such that the following square

$$\begin{array}{ccc} X_n & \xrightarrow{q_n} & X_{n+1} \\ \downarrow g_n & & \downarrow g_{n+1} \\ Y_n & \xrightarrow{p_n} & Y_{n+1} \end{array}$$

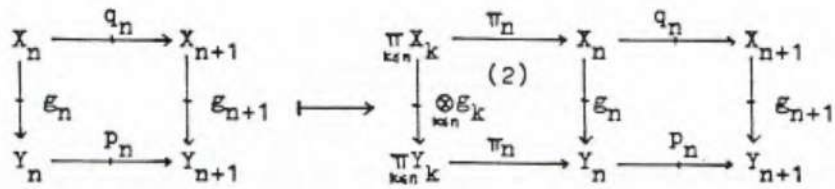
is commutative, i.e.  $g_{n+1} \cdot q_n = p_n \cdot g_n$ ,  $n = 0, 1, \dots$ . We denote this category as  $M_1$ .

Note that, the usual definition of Markov process is as above, but not as a sequence of probability measures on each state space  $X_n$ .

Definition 6. Define a functor  $F_1: S_1 \rightarrow M_1$ . For each object  $(X, q)$  in  $S_1$ , let  $F_1(X, q)$  be with state spaces  $1, X_1, X_1 \times X_2, \dots, \prod_{k=1}^n X_k, \dots$ , and dynamics  $(1_{\prod_{k=1}^n X_k}, q_n): \prod_{k=1}^n X_k \rightarrow \prod_{k=1}^{n+1} X_k$ . For each morphism  $f: (X, q) \rightarrow (Y, p)$  in  $S_1$ ,  
 defin  $F_1(f): F_1(X, q) \rightarrow F_1(Y, p)$   
 by  $F_1(f)_n = \otimes_{k=1}^n f_k: \prod_{k=1}^n X_k \rightarrow \prod_{k=1}^n Y_k$ .



Also define a functor  $G_1: M_1 \rightarrow S_1$ . For each object  $(X, q)$  in  $M_1$ , let  $G_1(X, q)$  be with state spaces  $1, X_1, \dots, X_n, \dots$ , and dynamics  $q_n \circ \pi_n: \prod_{k=1}^n X_k \rightarrow X_{n+1}$ . For each morphism  $g: (X, q) \rightarrow (Y, p)$  in  $M_1$ , Let  $G_1(g)_n = \varepsilon_n$ , for each  $n$ .



Then the square (1) is commutative, because the tensor in  $\mathcal{P}(m)$  is a bifunctor, this shows that  $F_1(f)$  is a morphism in  $M_1$ . The square (2) is commutative, because  $(\prod_{k=1}^n X_k, \pi_k, k=1, \dots, n)$  is a weak product in  $\mathcal{P}(m)$ , hence the rectangle on the right in above diagram is commutative, this shows that  $G_1(g)$  is a morphism in  $S_1$ . Thus,  $F_1$  and  $G_1$  are functors.

One of the difficulties here is that we can not develop any further relation between  $F_1$  and  $G_1$ .

We have more general definition of stochastic dynamics.

Definition 7. A stochastic dynamics is an object  $X = (X, q, h)$  in  $\mathcal{P}(M)^J$ , where  $(X, q, h) = (1 \xrightarrow{q_0} X_1 \xleftrightarrow{\dots} X_n \xleftrightarrow[h_n]{q_n} X_{n+1} \xleftrightarrow{\dots})$  satisfies  $h_n \circ q_n = 1_{X_n}$ , where  $X_n = X(n)$  and  $h_n$  is deterministic,  $n = 1, 2, \dots$ .

One example here is to take  $X_n$  as a subspace of the product space (accumulated state space until stage  $n$ ,  $\prod_{k=0}^n X_k$  as in definition 3) and  $h_n$  is the corresponding projection, for each  $n$ , such that  $h_n$ 's are compatible with  $X_n$ 's.

A morphism  $f: (X, q, h) \longrightarrow (Y, p, g)$  consists of a sequence of random maps  $f_n: X_n \longrightarrow Y_n$ , such that in the following diagram

$$\begin{array}{ccc}
 X_n & \xleftrightarrow[h_n]{q_n} & X_{n+1} \\
 \downarrow f_n & & \downarrow f_{n+1} \\
 Y_n & \xleftrightarrow[p_n]{g_n} & Y_{n+1}
 \end{array}$$

$p_n \circ f_n = f_{n+1} \circ q_n$ ,  $f_n \circ h_n = g_n \circ f_{n+1}$ . We denote this category as  $S_2$ .

The definition of the category of Markov dynamics is the same as in Definition 5, we denote it here as  $M_2$ .

Definition 8. Define  $F_2: S_2 \longrightarrow M_2$  as the forgetful functor, i.e. to forget the recall of history.

$$(X, q) \in M_2, (X, q) = (1 \xrightarrow{q_0} X_1 \xrightarrow{\dots} X_n \xrightarrow{q_n} X_{n+1} \xrightarrow{\dots})$$

define  $G_2: M_2 \longrightarrow S_2$ , by  $G_2(X, q)_0 = 1$ ,

$$G_2(X, q)_n = \prod_{k=1}^n X_k \xrightarrow{(1_{\prod X_k}, q_n \cdot \pi_n)} \prod_{k=1}^{n+1} X_k = G_2(X, q)_{n+1}, \quad n = 1, 2, \dots$$

If  $f: (X, q) \rightarrow (Y, p)$  is in  $M_2$ , a reasonable way to define

$G_2(f): G_2(X, q) \rightarrow G_2(Y, p)$  is as follows:  $G_2(f)_n = \otimes_{k=1}^n f_k$  for each  $n$ .

$$\begin{array}{ccc} \prod_{k=1}^n X_k & \xrightarrow{(1_{\prod X_k}, q_n \cdot \pi_n)} & \prod_{k=1}^{n+1} X_k \\ \downarrow \otimes_{k=1}^n f_k & & \downarrow \otimes_{k=1}^{n+1} f_k \\ \prod_{k=1}^n Y_k & \xrightarrow{(1_{\prod Y_k}, p_n \cdot \pi_n)} & \prod_{k=1}^{n+1} Y_k \end{array}$$

We would like that the above square commutes. As  $n=1$ , the above square

$$\text{is } \begin{array}{ccc} X_1 & \xrightarrow{(1_{X_1}, q_1)} & X_1 \times X_2 \\ \downarrow f_1 & & \downarrow f_1 \otimes f_2 \\ Y_1 & \xrightarrow{(1_{Y_1}, p_1)} & Y_1 \times Y_2 \end{array} = \begin{array}{ccccc} X_1 & \xrightarrow{\Delta_{X_1}} & X_1 \times X_1 & \xrightarrow{1_{X_1} \otimes q_1} & X_1 \times X_2 \\ \downarrow f_1 & (1) & \downarrow f_1 \otimes f_1 & (2) & \downarrow f_1 \otimes f_2 \\ Y_1 & \xrightarrow{\Delta_{Y_1}} & Y_1 \times Y_1 & \xrightarrow{1_{Y_1} \otimes p_1} & Y_1 \times Y_2 \end{array}$$

the square (2) is commutative, because  $\otimes$  is a bifunctor and  $f_2 \circ q_1 = p_1 \circ f_1$ ; but (1) may not be commutative in  $\mathcal{P}(m)$ , because  $\Delta$  is not natural in  $\mathcal{P}(m)$ .

$\Delta$  is not natural and  $(\prod_{k=1}^n X_k, \pi_k, k=1, \dots, n)$  is not the product but only a weak product of  $X_k$ 's in  $\mathcal{P}(m)$ , these are the reasons that we could not show that  $G_2$  is a functor and develop any further relation between  $F_1$  and  $G_1$ .

There are two ways to overcome these difficulties, one is to

"shrink"  $S_2, M_2$  by taking only deterministic maps as dynamics and morphisms between stochastic dynamics; another way is to enlarge  $S_2$  and  $M_2$ .

Let  $S_0, M_0$  be corresponding deterministic subcategories (i.e. dynamics and morphisms consist only deterministic maps) of  $S_2, M_2$ ;  $F_0, G_0$  be the corresponding restrictions of  $F_2, G_2$  on  $S_0, M_0$ .

Theorem.  $F_0$  is left adjoint to  $G_0$ .

Proof. Let  $(X, q, h) = (1 \xrightarrow{q_0} X_1 \xleftrightarrow{\quad} \dots \xleftrightarrow{\quad} X_n \xleftrightarrow{h_n} X_{n+1} \xleftrightarrow{q_n} \dots) \in S_0$ .

Define  $\eta_X: (X, q, h) \longrightarrow G_0 F_0 (X, q, h)$  by

$$\begin{array}{ccc} X_{n+1} & \xrightarrow{(\eta_X)_{n+1}} & \prod_{k \geq n+1} X_k \\ \downarrow h_n & & \downarrow \pi_{(n)} \\ X_n & \xrightarrow{(\eta_X)_n} & \prod_{k \geq n} X_k \end{array}$$

$$(\eta_X)_n = (h_1 \circ \dots \circ h_{n-1}, \dots, h_{n-1}, 1_{X_n}),$$

$$(\eta_X)_n \circ h_n = (h_1 \circ \dots \circ h_{n-1}, \dots, h_{n-1}, 1_{X_n}) \circ h_n$$

$$= (h_1 \circ \dots \circ h_{n-1} \circ h_n, \dots, h_{n-1} \circ h_n, h_n)$$

$$= \pi_{(n)}(h_1 \circ \dots \circ h_n, \dots, h_{n-1} \circ h_n, h_n, 1_{X_{n+1}}) = \pi_{(n)}(\eta_X)_{n+1}$$

this shows that  $\eta_X$  is a morphism in  $S_0$ .

If  $f: (X, q, h) \longrightarrow (Y, p, g)$  is a morphism in  $S_0$ , i.e.

$f_{n+1} \circ q_n = p_n \circ f_n$  and  $f_n \circ h_n = g_n \circ f_{n+1}$ ,  $n = 0, 1, \dots$  (see the diagram in p.3.20). Then in the following diagram

$$\begin{array}{ccc}
 X_n & \xrightarrow{(\gamma_X)_n} & \prod_{k \in \mathbb{N}} X_k \\
 \downarrow f_n & & \downarrow \prod_{k \in \mathbb{N}} f_k \\
 Y_n & \xrightarrow{(\gamma_Y)_n} & \prod_{k \in \mathbb{N}} Y_k
 \end{array}$$

$$\begin{aligned}
 (\gamma_Y)_n \circ f_n &= (g_1 \circ \dots \circ g_{n-1} \circ \dots \circ g_{n-1} \circ 1_{Y_n}) \circ f_n \\
 &= (g_1 \circ \dots \circ g_{n-1} \circ f_n, \dots, g_{n-1} \circ f_n, f_n)
 \end{aligned}$$

$$\begin{aligned}
 \text{and } \prod_{k \in \mathbb{N}} f_k \circ (\gamma_X)_n &= \prod_{k \in \mathbb{N}} f_k \circ (h_1 \circ \dots \circ h_{n-1} \circ \dots \circ h_{n-1} \circ 1_{X_n}) \\
 &= (f_1 \circ h_1 \circ \dots \circ h_{n-1}, \dots, f_{n-1} \circ h_{n-1}, f_n)
 \end{aligned}$$

$$\begin{aligned}
 \text{but } g_1 \circ \dots \circ g_{n-1} \circ f_n &= g_1 \circ \dots \circ f_{n-1} \circ h_{n-1} \quad (\text{use } f_n \circ h_n = g_n \circ f_{n+1}) \\
 &= g_1 \circ \dots \circ f_{n-2} \circ h_{n-2} \circ h_{n-1} \\
 &\quad \dots \\
 &= f_1 \circ h_1 \circ \dots \circ h_{n-1} \\
 &\quad \dots
 \end{aligned}$$

$$g_{n-1} \circ f_n = f_{n-1} \circ h_{n-1}$$

thus,  $(\gamma_Y)_n \circ f_n = \prod_{k \in \mathbb{N}} f_k \circ (\gamma_X)_n$ , for all  $n$ . This shows that  $\gamma$  is natural as the unit of  $F_0$  and  $G_0$ .

If  $(Y, p) \in M_0$  and  $\varphi: (X, q, h) \longrightarrow G_0(Y, p)$  is a morphism in  $S_0$ , i.e. for each  $n$ ,  $\varphi_n: X_n \longrightarrow \prod_{k \in \mathbb{N}} Y_k$  is a measurable function and in the following sequence of squares

$$\begin{array}{ccccc}
 1 & \xrightarrow{q_0} & X_1 & \xrightleftharpoons[h_1]{q_1} & X_2 & \cdots & X_n & \xrightleftharpoons[h_n]{q_n} & X_{n+1} \\
 & \searrow p_0 & \downarrow \varphi_1 & & \downarrow \varphi_2 & \cdots & \downarrow \varphi_n & & \downarrow \varphi_{n+1} & \cdots \\
 & & Y_1 & \xrightleftharpoons[\pi_1]{(1_{Y_1}, p_1)} & Y_1 \times Y_2 & & \prod_{k \in \mathbb{N}} Y_k & \xrightleftharpoons[\pi_{(n)}]{(1_{\prod_{k \in \mathbb{N}} Y_k}, p_n \circ \pi_n)} & \prod_{k \in \mathbb{N}+1} Y_k &
 \end{array}$$

$$\varphi_1 \circ q_0 = p_0, \dots, \varphi_{n+1} \circ q_n = (1_{\prod_{k \in \mathbb{N}} Y_k}, p_n \circ \pi_n) \circ \varphi_n, \varphi_n \circ h_n = \pi_{(n)} \circ \varphi_{n+1},$$

for  $n = 1, 2, \dots$

Define  $\bar{\varphi} : F_0(X, q, h) \longrightarrow (Y, p)$ , by

$$\begin{aligned}\bar{\varphi}_n &= \pi_n \circ \varphi_n : X_n \longrightarrow \prod_{k \leq n} Y_k \longrightarrow Y_n, \\ \text{then } \bar{\varphi}_{n+1} \circ q_n &= \pi_{n+1} \circ \varphi_{n+1} \circ q_n = \pi_{n+1} \circ (1_{\prod_{k \leq n} Y_k}, p_n \circ \pi_n) \circ \varphi_n \\ &= \pi_{n+1} \circ (\varphi_n, p_n \circ \pi_n \circ \varphi_n) = p_n \circ \pi_n \circ \varphi_n = p_n \circ \bar{\varphi}_n\end{aligned}$$

thus  $\bar{\varphi}$  is a morphism in  $M_0$ .

Note that, if let  $\pi_n^m : \prod_{i \leq m} Y_i \longrightarrow Y_n$ ,  $n \leq m$ , be the projection,

$$\begin{aligned}\text{then } \bar{\varphi}_k \circ h_k \circ h_{k+1} \circ \dots \circ h_{n-1} &= \pi_k^k \circ (\varphi_k \circ h_k) \circ \dots \circ h_{n-1} \\ &= \pi_k^k \circ \pi_{(k)} \circ \varphi_{k+1} \circ h_{k+1} \circ \dots \circ h_{n-1} \\ &= \pi_k^{k+1} \circ \varphi_{k+1} \circ h_{k+1} \circ \dots \circ h_{n-1} \\ &= \pi_k^{k+1} \circ \pi_{(k+1)} \circ \varphi_{k+2} \circ \dots \circ h_{n-1} \\ &\dots \\ &= \pi_k^n \circ \varphi_n.\end{aligned}$$

Claim.  $(G_0 \bar{\varphi}) \circ \gamma_X = \varphi$ .

$$\begin{aligned}(G_0 \bar{\varphi})_n \circ (\gamma_X)_n &= \prod_{k \leq n} \bar{\varphi}_k \circ (h_1 \circ \dots \circ h_{n-1}, \dots, h_{n-1}, 1_{X_n}) \\ &= (\bar{\varphi}_1 \circ h_1 \circ \dots \circ h_{n-1}, \dots, \bar{\varphi}_{n-1} \circ h_{n-1}, \bar{\varphi}_n) \\ &= (\pi_1^n \circ \varphi_n, \dots, \pi_{n-1}^n \circ \varphi_n, \pi_n^n \circ \varphi_n) \\ &= \varphi_n.\end{aligned}$$

$\bar{\varphi}$  is unique: if  $\varphi' : F_0(X, q, h) \longrightarrow (Y, p)$  is another morphism such that  $G_0 \varphi' \circ \gamma_X = \varphi$ . Then for any  $n$ ,

$$\varphi_n = (G_0 \varphi')_n \circ (\gamma_X)_n = \prod_{k \leq n} \varphi'_k \circ (h_1 \circ \dots \circ h_{n-1}, \dots, h_{n-1}, 1_{X_n})$$

$$= (\varphi'_1 \cdot h_1 \cdot \dots \cdot h_{n-1}, \dots, \varphi'_{n-1} \cdot h_{n-1}, \varphi'_n)$$

so,  $\pi_n \cdot \varphi_n = \varphi'_n$ , but  $\pi_n \cdot \varphi_n = \bar{\varphi}_n$ . Thus,  $\varphi'_n = \bar{\varphi}_n$ .

This completes the proof of the theorem.

Remark. Let  $\omega h$  be the category

$$0 \begin{array}{c} \xrightarrow{d_{01}} \\ \xleftarrow{h_{10}} \end{array} 1 \begin{array}{c} \xrightarrow{d_{12}} \\ \xleftarrow{h_{21}} \end{array} 2 \longleftrightarrow \dots \longleftrightarrow n \begin{array}{c} \xrightarrow{d_{n,n+1}} \\ \xleftarrow{h_{n+1,n}} \end{array} n+1 \longleftrightarrow \dots$$

such that  $h_{n+1,n} \cdot d_{n,n+1} = 1_n$ .

Note,  $\omega h(n,m) = \begin{cases} \{d_{0m} \cdot h_{n0}, d_{1m} \cdot h_{n1}, \dots, d_{n-1,m} \cdot h_{n,n-1}, d_{nm}\}, & n \leq m, \\ \{d_{0m} \cdot h_{n0}, d_{1m} \cdot h_{n1}, \dots, d_{m-1,m} \cdot h_{n,m-1}, h_{nm}\}, & n > m \end{cases}$

and  $\omega h(n,m)$  has  $\begin{cases} n+1 \text{ elements, for } n \leq m, \\ m+1 \text{ elements, for } n > m. \end{cases}$

Let  $\omega$  be the category

$$0 \xrightarrow{e_{01}} 1 \xrightarrow{e_{12}} 2 \longrightarrow \dots \longrightarrow n \xrightarrow{e_{n,n+1}} n+1 \longrightarrow \dots$$

$\mathcal{M}$  is the category of measurable spaces and measurable functions.

Let  $\mathcal{M}^{\omega h}$  be the category of functors from  $\omega h$  to  $\mathcal{M}$ ,  $\mathcal{M}^{\omega}$  be the category of functors from  $\omega$  to  $\mathcal{M}$ . Then  $S_0 \subset \mathcal{M}^{\omega h}$  and  $M_0 \subset \mathcal{M}^{\omega}$ .

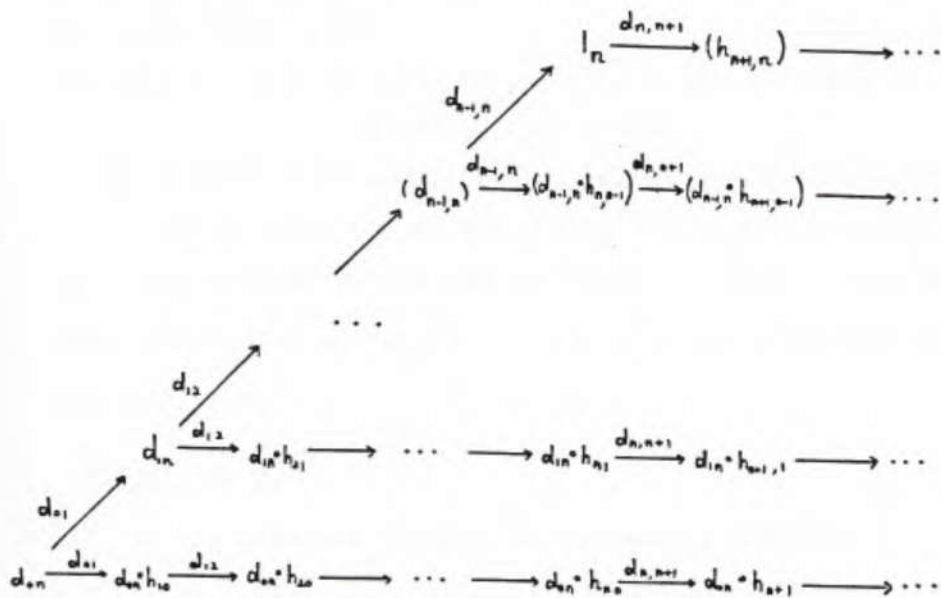
((X,q,h)  $\in S_0$  needs  $X_0 = 1$ ; similarly (Y,p)  $\in M_0$ ,  $Y_0 = 1$ ).

The inclusion functor  $i: \omega \hookrightarrow \omega h$  induces the forgetful functor  $F: \mathcal{M}^{\omega h} \rightarrow \mathcal{M}^{\omega}$ .  $i$  has the right Kan extension  $G: \mathcal{M}^{\omega} \rightarrow \mathcal{M}^{\omega h}$ , which is the right adjoint of  $F$ . The restrictions of  $F$  and  $G$  on  $S_0$  and  $M_0$  are just  $F_0$  and  $G_0$ .

$i$  also has the left Kan extension  $\bar{G}$ , since  $\mathcal{M}$  is also cocomplete.

$X \in \mathcal{M}^{\omega}$ , for each  $n$ ,  $(\bar{G}X)_n$  is the colimit of  $(i|n) \xrightarrow{P} \omega \xrightarrow{X} m$ ,  
 where  $P$  is the projection,  $P$  sends  $k \xrightarrow{\quad} m$  in  $(i|n)$  to  $k \rightarrow m$ ,

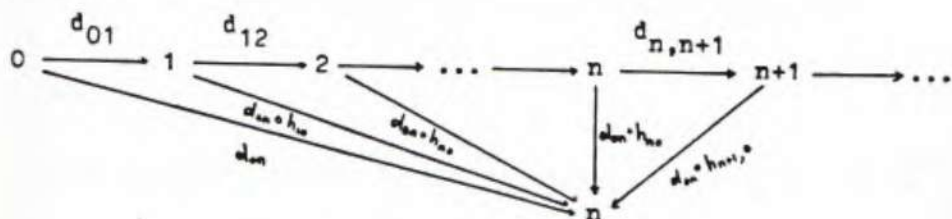
i.e.  $(\bar{G}X)_n$  is the colimit of  $X$  acting on the following diagram



where

$$d_{0,n} \xrightarrow{d_{0,1}} d_{0,n} * h_{n,n} \xrightarrow{d_{1,2}} d_{0,n+1} * h_{n,n} \xrightarrow{\dots} d_{0,n+1} * h_{n,n+1} \xrightarrow{d_{n,n+1}} d_{0,n+1} * h_{n,n+1} \xrightarrow{\dots}$$

means



On the other hand, we could extend  $\mathcal{P}(M)$  to  $\mathcal{C}$ .

Lemma. There is a faithful functor  $\alpha: \mathcal{P}(M) \longrightarrow \mathcal{C}$ .

Proof. For  $X \in \mathcal{P}(M)$ , define  $\alpha(X) = \mathcal{P}(X)$  the probability measures on  $X$ . If  $f: X \longrightarrow Y$  is a random map, define  $\alpha(f): \mathcal{P}(X) \longrightarrow \mathcal{P}(Y)$  by  $p \longmapsto f \cdot p$ ,  $p \in \mathcal{P}(X)$ . Then  $\alpha(f)$  is affine linear, because for  $p, q \in \mathcal{P}(X)$ ,  $t \in [0, 1]$ ,

$$\begin{aligned} \alpha(f)(tp + (1-t)q) &= f \cdot (tp + (1-t)q) = tf \cdot p + (1-t)f \cdot q \\ &= t\alpha(f)(p) + (1-t)\alpha(f)(q). \end{aligned}$$

$\alpha$  is faithful: if  $f$  and  $g$  are two random maps from  $X$  to  $Y$ ,  $f \neq g$  i.e. there exists  $x \in X$  and  $B \in \mathcal{B}$ , such that  $f(x, B) \neq g(x, B)$ . Let  $\delta_x$  be the probability measure on  $X$  which is concentrated on  $x$ , then  $\alpha(f)(\delta_x)(B) = f \cdot \delta_x(B) = f(x, B) \neq g(x, B) = g \cdot \delta_x(B) = \alpha(g)(\delta_x)(B)$ , i.e.  $\alpha(f) \neq \alpha(g)$ .

Definition 9.

$\mathcal{C}^{wh}$  is the extended category of stochastic dynamics,

$\mathcal{C}^w$  is the extended category of Markov dynamics,

$\mathcal{C}^2$  is the extended category of stationary Markov dynamics, where

$\mathcal{C}^2 = \mathcal{C}^N$ ,  $N$  is the additive monoid of natural numbers  $\{0, 1, \dots\}$ .

Remark. a). From Definition 7, a stochastic dynamics is

$$(X, q, h) = (1 \xrightarrow{q_0} X_1 \xrightleftharpoons[h_1]{q_1} X_2 \xrightarrow{\dots} X_n \xrightleftharpoons[h_n]{q_n} X_{n+1} \xrightarrow{\dots})$$

with  $h_n \cdot q_n = 1_{X_n}$ ,  $n = 1, 2, \dots$ ,

using the above Lemma, we get an object  $(\mathcal{P}(X_n), \bar{q}_n, \bar{h}_n)$  in  $\mathcal{C}^{wh}$ :

$$1 \xrightarrow{\bar{q}_0} \mathcal{P}(X_1) \xrightleftharpoons[\bar{h}_1]{\bar{q}_1} \mathcal{P}(X_2) \rightleftarrows \dots \rightleftarrows \mathcal{P}(X_n) \xrightleftharpoons[\bar{h}_n]{\bar{q}_n} \mathcal{P}(X_{n+1}) \rightleftarrows \dots$$

with  $\bar{h}_n \circ \bar{q}_n = \text{id}_{\mathcal{P}(X_n)}$ ,  $n = 1, 2, \dots$ .

b). In the above definitions, we made stochastic dynamics  $\mathcal{C}^{wh}$  and Markov dynamics  $\mathcal{C}^w$  more general, namely instead of an initial probability measure, we use an initial random map.

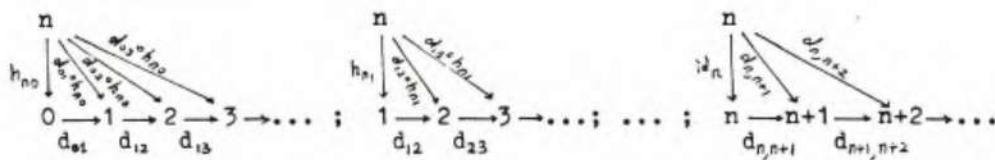
Theorem 2. The inclusion functor  $K: \mathcal{W} \hookrightarrow \mathcal{Wh}$  induces the forgetful functor  $F: \mathcal{C}^{Wh} \longrightarrow \mathcal{C}^w$ .  $K$  has the right Kan extension  $G: \mathcal{C}^w \longrightarrow \mathcal{C}^{Wh}$ , which sends

$$(A, f) = (A_0 \xrightarrow{f_{01}} A_1 \xrightarrow{f_{12}} A_2 \longrightarrow \dots \longrightarrow A_n \xrightarrow{f_{n,n+1}} A_{n+1} \longrightarrow \dots) \text{ in } \mathcal{C}^w$$

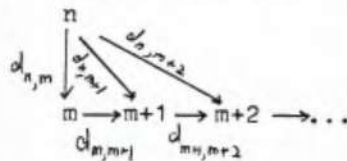
$$\text{to } G(A, f) = (A_0 \xleftarrow[\pi_0]{(1_{A_0}, f_{01})} A_0 \times A_1 \xleftarrow[\pi_{12}]{(1_{A_1}, f_{12}, \pi_2)} \prod A_k \rightleftarrows \dots \rightleftarrows \prod A_k \xleftarrow[\pi_{(n)}]{(1_{\prod A_k}, f_{n,n+1}, \pi_n)} \prod A_k \rightleftarrows \dots)$$

in  $\mathcal{C}^{Wh}$ . So,  $G$  is the right adjoint of  $F$ .  $K$  also has the left adjoint  $\bar{G}$ .

Proof. For each  $n \in \mathcal{Wh}$ ,  $(n|K)$  consists of strings of following types:



each triangle is commutative. Note that, for  $m > n$ , the string

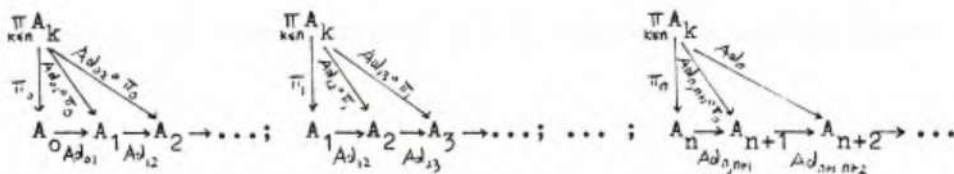


is just a part of the last one in the above.

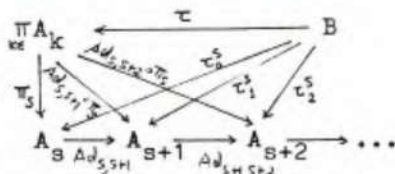
$A \in \mathcal{C}^\omega$ ,  $n \in \omega$ , the right Kan extension  $(GA)_n$  is the limit of

$$(n|K) \xrightarrow{Q_n} \omega \xrightarrow{A} \mathcal{C}, \text{ where } Q_n \text{ is the forgetful functor.}$$

Since for each  $n$ , there are  $n+1$  different types of strings in  $(n|K)$ , therefore  $\varprojlim AQ_n = \prod_{k \in n} A_k$  with the projections



if  $B$  is an object in  $\mathcal{C}$ , such that for each string  $s \in n$ ,



$$Ad_{s,s+1} \tau_s = \tau_{s+1}, \quad Ad_{s+1,s+2} \tau_{s+1} = \tau_{s+2}, \dots \quad \text{Define}$$

$$\tau = (\tau_0, \tau_1, \dots, \tau_n): B \longrightarrow \prod_{k \in n} A_k, \quad \text{then } \pi_s \cdot \tau = \tau_s, \text{ for } s \in n.$$

The morphisms  $n \xrightarrow{d_{n,n+1}} n+1$ ,  $n = 0, 1, 2, \dots$ , induce maps  $n+1, n$

$$(GA)_n = \varprojlim AQ_n \xleftarrow[\pi_{(n)}]{(1_{\prod_{k \in n} A_k}, Ad_{n,n+1} \cdot \pi_n)} (GA)_{n+1} = \varprojlim AQ_{n+1}.$$

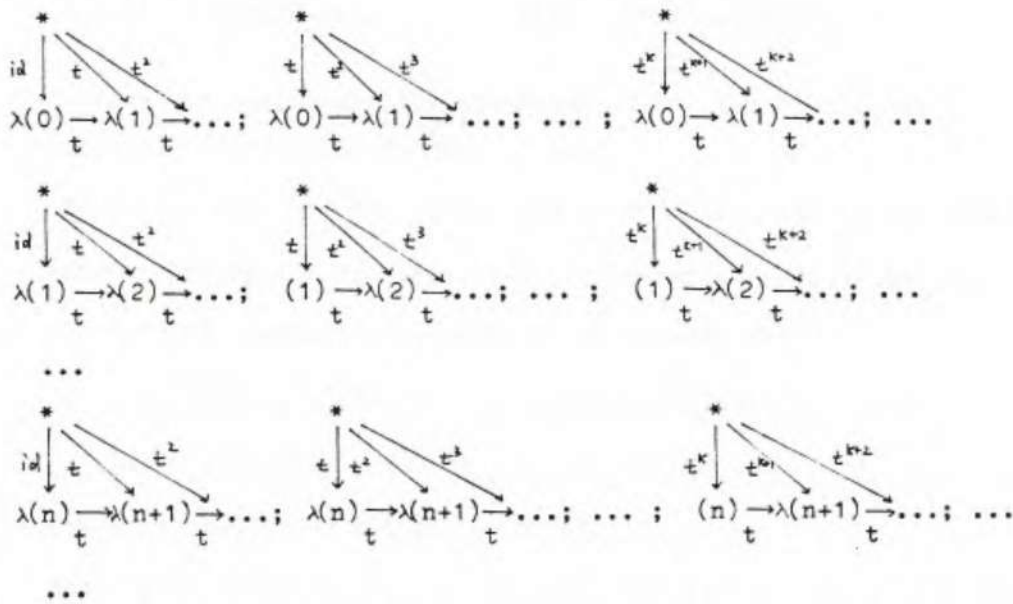
The left Kan extension  $\bar{G}$  of the inclusion functor  $K$  is constructed in a similar way as in Theorem 1.

**Definition 10.** Define a functor  $\lambda : \omega \longrightarrow N = \mathcal{D}^\dagger$ , by  $\lambda(n) = *$ ,  $*$  is the only object in  $N$  and  $\lambda(d_{n,n+1}) = t$ , for all  $n$ .

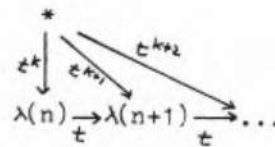
$\lambda$  induces an inclusion functor  $L: \mathcal{C}^2 \longrightarrow \mathcal{C}^{\omega}$ , for  $X \in \mathcal{C}^2$ ,  
 $(LX)_n = X(*)$  and  $(LX)(d_{n,n+1}) = X(t)$ .

Theorem 3. The inclusion functor  $L: \mathcal{C}^2 \longrightarrow \mathcal{C}^{\omega}$  has right adjoint  $G$  and left adjoint  $\bar{G}$ , namely the right and left Kan extensions of  $\lambda: \omega \longrightarrow \mathcal{C}^t$ , such that for  $B \in \mathcal{C}^{\omega}$ ,  $GB = \prod_{k=1}^{\infty} B_k$ ,  $\bar{G}B = \prod_{k=0}^{\infty} B_k$ .

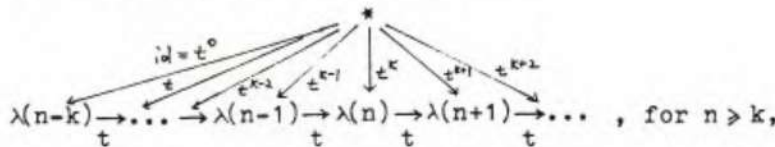
Proof. The comma category  $(*\downarrow \lambda)$  consists of strings like follows:



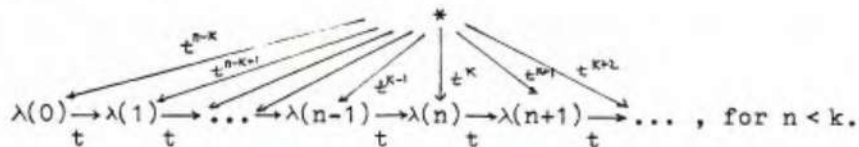
But each string of the type



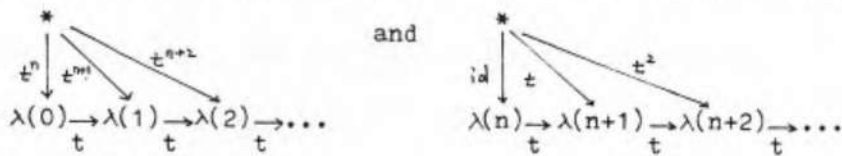
is either as a part of the string



or as a part of the string



So, for each n, there are two different types of strings



Therefore, the right Kan extension of  $\lambda$ , for  $B \in \mathcal{C}^{\omega}$ , the limit of  $(*|\lambda) \xrightarrow{Q} \omega \xrightarrow{B} \mathcal{C}$ , is

$GB = \varprojlim BQ = \prod_0 \times \prod_{k=1}^{\infty} B_k$ , for each n,  $B_0$  takes care of the first type of string and  $B_n$  takes care of the second type of string.

$t: * \rightarrow *$  induces a morphism  $g: GB \rightarrow GB$ , say

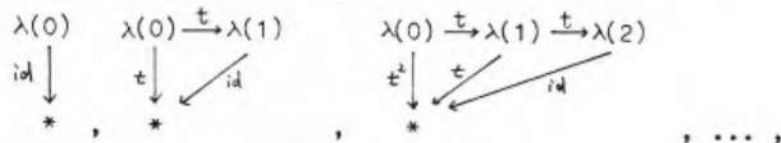
$$B = (B_0 \xrightarrow{g_{01}} B_1 \xrightarrow{g_{12}} B_2 \rightarrow \dots \rightarrow B_n \xrightarrow{g_{n,n+1}} B_{n+1} \rightarrow \dots),$$

let  $b \in GB = \prod_0 \times \prod_{k=1}^{\infty} B_k$ ,  $b = (b_{00}, b_{01}, b_{02}, \dots; b_1, b_2, \dots, b_n, \dots)$

where  $b_{0k} \in B_0$ ,  $k = 0, 1, 2, \dots$ ;  $b_i \in B_i$ ,  $i \geq 1$ ,

then  $g(b) = (b_{01}, b_{02}, b_{03}, \dots; g_{01}(b_{00}), g_{12}(b_1), \dots, g_{n-1,n}(b_{n-1}), \dots)$ .

For the left Kan extension  $\bar{U}: \mathcal{C}^{\omega} \rightarrow \mathcal{C}^{\omega}$ , the comma category  $(\lambda|*)$  consists of finite length strings:



$$\begin{array}{ccccccc}
 \lambda(0) & \xrightarrow{t} & \lambda(1) & \xrightarrow{t} & \lambda(2) & \longrightarrow & \dots \longrightarrow \lambda(n-2) & \xrightarrow{t} & \lambda(n-1) & \xrightarrow{t} & \lambda(n) \\
 \downarrow t^n & \nearrow t^{n-1} & \nearrow t^{n-2} & \nearrow t^{n-3} & \nearrow t^{n-4} & \nearrow t^{n-5} & \nearrow t^{n-6} & \nearrow t^{n-7} & \nearrow t^{n-8} & \nearrow t^{n-9} & \nearrow \text{id} \\
 * & & & & & & & & & & 
 \end{array}$$

If  $R \in \mathcal{C}^w$ ,  $B = (B_0 \xrightarrow{\varepsilon_{01}} B_1 \xrightarrow{\varepsilon_{12}} B_2 \longrightarrow \dots \longrightarrow B_n \xrightarrow{\varepsilon_{n,n+1}} B_{n+1} \longrightarrow \dots)$ ,

then  $\bar{B} = \prod_{k=0}^{\infty} B_k$ , and  $\bar{g} : \prod_{k=0}^{\infty} B_k \longrightarrow \prod_{k=0}^{\infty} B_k$  is induced by

$$\varepsilon_{k,k+1} : B_k \longrightarrow B_{k+1}, \quad k = 0, 1, 2, \dots$$

### §3. Decomposition of Stochastic Transformations.

In Heller's paper [6], the category of stochastic transformations consists of only those  $X$ 's, such that  $X_n = X_0^n$ ,  $X_0$  is a finite set,  $n = 1, 2, \dots$  and the recall of history is just the projection. For each such  $X$ , let  $A(X_0)$  be the free associative  $R$ -algebra generated by  $X_0$ , where  $R$  is the reals, then there is a  $A(X_0 \times Y_0)$ -module structure associated with every stochastic transformation  $\varphi$  from  $X$  to  $Y$ . He applies the decomposition theorems of modules to stochastic transformations.

Here, we discuss the decomposition problem of stochastic transformations in a direct way and also in a larger category.

Definition 1.  $S'$  is a full sub category of the category of stochastic transformations  $\mathcal{O}(m^{\omega})$  with objects  $X$ , such that  $X_n$  are finite sets and  $h_n: X_{n+1} \rightarrow X_n$  are surjective,  $n = 0, 1, \dots$ ;  $S''$  is a full sub category of  $S'$ , such that  $X$  is an object in  $S''$ , then  $X_n = X_0^{n+1}$ ,  $X_0$  is a finite set and  $h_n: X_0^{n+2} \rightarrow X_0^{n+1}$  are just the projections,  $n = 0, 1, \dots$ .

Notations. 1).  $\pi_{(n)}$  is the projection:  $\prod_{k=1}^{n+1} X_k \rightarrow \prod_{k=1}^n X_k$ .

2).  $\varphi: X \rightarrow Y$  is a stochastic transformation, then use  $\varphi_k$  to denote the  $k$ -th component of  $\varphi$ ,  $\varphi_k: X_k \rightarrow Y_k$ .

Lemma 1.  $\varphi: X \rightarrow Y$  is a random map,  $X$  and  $Y$  are finite sets. Then there exist  $t_i \geq 0$ ,  $\sum_{i=1}^n t_i = 1$  and deterministic maps

$\varphi_i: X \rightarrow Y$ ,  $i = 1, 2, \dots, n$ , such that  $\varphi = \sum_{i=1}^n t_i \varphi_i$ .

Proof. Let  $t_1 = \min_{\varphi(x,y) \neq 0} \varphi(x,y)$ , say  $t_1 = \varphi(x_1, y_1)$ . Define  $\varphi_1: X \rightarrow Y$  as follows: if  $x = x_1$ ,  $\varphi_1(x_1) = y_1$ ;

if  $x \neq x_1$ ,  $\varphi_1(x) = y_x$ , for some  $y_x \in Y$ , such that  $\varphi(x, y_x) \neq 0$ .

We could view  $\varphi_1$  as a random map in the following way:

$$\varphi_1(x, y) = \begin{cases} 1, & y = y_1, \text{ if } x = x_1 \text{ or } y = y_x, \text{ if } x \neq x_1, \\ 0, & \text{otherwise.} \end{cases}$$

Define  $\varphi^{(1)}: X \rightarrow Y$ ,

$$\varphi^{(1)}(x, y) = \frac{1}{1-t_1} (\varphi(x, y) - t_1 \varphi_1(x, y)).$$

If  $\varphi(x, y) \neq 0$  then  $\varphi(x, y) \geq t_1$  and  $0 \leq \varphi_1(x, y) \leq 1$ ; if  $\varphi(x, y) = 0$ , this implies  $\varphi_1(x, y) = 0$ , so  $\varphi^{(1)}(x, y) \geq 0$  in both cases.  $\sum_{y \in Y} \varphi^{(1)}(x, y) = \frac{1}{1-t_1} (\sum_{y \in Y} \varphi(x, y) - t_1 \sum_{y \in Y} \varphi_1(x, y)) = \frac{1}{1-t_1} (1-t_1) = 1$ , so  $\varphi^{(1)}(x, y) \leq 1$ , i.e.  $\varphi^{(1)} \in \mathcal{P}(m)(X, Y)$  and it is easy to see that  $\varphi = t_1 \varphi_1 + (1-t_1) \varphi^{(1)}$ .

Repeating same procedure to  $\varphi^{(1)}$ , we get  $\varphi^{(1)} = s_2 \varphi_2 + (1-s_2) \varphi^{(2)}$ , where  $\varphi_2$  is deterministic, and  $\varphi = t_1 \varphi_1 + (1-t_1)(s_2 \varphi_2 + (1-s_2) \varphi^{(2)})$ , let  $t_2 = (1-t_1)s_2$ , then  $\varphi = t_1 \varphi_1 + t_2 \varphi_2 + (1-t_1-t_2) \varphi^{(2)}$ . Repeating the same procedure to  $\varphi^{(2)}, \dots$ , since  $X, Y$  are finite, eventually we get  $\varphi = \sum_{i=1}^n t_i \varphi_i$ , where  $\varphi_i$ ,  $i = 1, \dots, n$ , are deterministic.

Remark. a). This decomposition of  $\varphi$  is far from unique.  
b). Let  $Y^X$  be the set of arbitrary maps from  $X$  to  $Y$ . Since  $X, Y$  are finite discrete spaces, any map from  $X$  to  $Y$  is a measurable function, i.e. it could be viewed as a random map, hence there is an inclusion  $Y^X \rightarrow \mathcal{P}(m)(X, Y)$ . For any set  $V$ ,  $\mathcal{P}(m)(1, V)$  is the free convex set generated by  $V$ , specially  $\mathcal{P}(m)(1, Y^X)$  is the free convex set generated by  $Y^X$ , so there is a unique affine

linear map,  $\mathcal{P}(m)(1, Y^X) \rightarrow \mathcal{P}(m)(X, Y)$ . Lemma 1 says that this map is surjective.

Lemma 2.  $X, Y, n$  and  $m$  are finite sets,  $f: X \rightarrow n$ ,  $g: Y \rightarrow m$  are surjective deterministic maps. If the square

$$\begin{array}{ccc} X & \xrightarrow{\varphi} & Y \\ \downarrow f & & \downarrow g \\ n & \xrightarrow{\psi} & m \end{array}$$

is commutative, then  $\varphi$  is deterministic implies  $\psi$  is deterministic.

Proof. Since  $f: X \rightarrow n$  is surjective, so there is a right inverse  $h: n \rightarrow X$  of  $f$ , i.e.  $f \cdot h = 1_n$ . If  $\varphi$  is deterministic, then  $\psi = \psi \cdot f \cdot h = g \cdot \varphi \cdot h$  is deterministic.

Lemma 3. In the following diagram

$$\begin{array}{ccccccc} X_n & \xrightarrow{f_{n-1}} & \dots & \xrightarrow{f_k} & X_{k-1} & \xrightarrow{\dots} & \xrightarrow{f_0} & X_0 \\ \downarrow \varphi_n & & & \downarrow \varphi_k & \downarrow \varphi_{k-1} & & \downarrow \varphi_0 & \\ Y_n & \xrightarrow{g_{n-1}} & \dots & \xrightarrow{g_{k-1}} & Y_{k-1} & \xrightarrow{\dots} & \xrightarrow{g_0} & Y_0 \end{array}$$

$\varphi_k$  are random maps,  $k = 0, 1, \dots, n$ ;  $f_k, g_k$  are surjective deterministic maps,  $k = 0, 1, \dots, n-1$ , such that

$\varphi_{k-1} \circ f_{k-1} = g_{k-1} \circ \varphi_k$ . Then, we can decompose  $\varphi_k$ 's in such a way,  $\varphi_k = \sum_{i=1}^s t_i \varphi_{ki}$ ,  $t_i \geq 0$ ,  $\sum_{i=1}^s t_i = 1$ ,  $\varphi_{ki}$  are deterministic,  $k = 0, 1, \dots, n$  and  $\varphi_{k-1, i} \circ f_{k-1} = g_{k-1} \circ \varphi_{ki}$ , for  $i = 1, \dots, s$ .

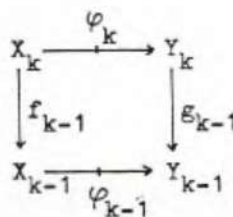
Proof. Denote  $f_k^m = f_k \circ \dots \circ f_{m-1}$ ,  $g_k^m = g_k \circ \dots \circ g_{m-1}$ ,

for  $0 \leq k < m-1 \leq n-1$ .

$$\varphi_{k-1} \circ f_{k-1} = g_{k-1} \circ \varphi_k, \quad k = 1, \dots, n,$$

i.e. for any  $x_k \in X_k$ ,  $y_{k-1} \in Y_{k-1}$ ,

$$\begin{aligned} & \varphi_{k-1}(f_{k-1}(x_k), y_{k-1}) \\ &= \varphi_{k-1} \circ f_{k-1}(x_k, y_{k-1}) = g_{k-1} \circ \varphi_k(x_k, y_{k-1}) \\ &= \varphi_k(x_k, g_{k-1}^{-1}(y_{k-1})). \end{aligned}$$



Let  $t_1 = \min_{\varphi_n(x_n, y_n) \neq 0} \varphi_n(x_n, y_n)$ , say  $t_1 = \varphi_n(x_n^0, y_n^0)$ .

Define  $\varphi_{01}: X_0 \rightarrow Y_0$ ,

if  $x_0 = f_0^n(x_n^0)$ ,  $\varphi_{01}(x_0) = g_0^n(y_n^0)$ ;

if  $x_0 \neq f_0^n(x_n^0)$ ,  $\varphi_{01}(x_0) = y_{x_0}$ , for some  $y_{x_0} \in Y$ , such that

$$\varphi_0(x_0, y_{x_0}) \neq 0.$$

So,  $\varphi_0(x_0, \varphi_{01}(x_0)) \neq 0$ .

Define  $\varphi_{11}: X_1 \rightarrow Y_1$ ,

if  $x_1 = f_1^n(x_n^0)$ ,  $\varphi_{11}(x_1) = g_1^n(y_n^0)$ ;

if  $x_1 \neq f_1^n(x_n^0)$ ,  $\varphi_{11}(x_1) = y_{x_1}$ , where  $y_{x_1} \in g_0^{-1}(\varphi_{01}(f_0(x_1)))$ ,

$$\text{such that } \varphi_1(x_1, y_{x_1}) \neq 0.$$

(such  $y_{x_1}$  exists, because  $\varphi_1(x_1, g_0^{-1}(\varphi_{01}(f_0(x_1))))$

$$= \varphi_0(f_0(x_1), \varphi_{01}(f_0(x_1))) \neq 0, \text{ so } \varphi_1(x_1, \varphi_{11}(x_1)) \neq 0.$$

Assume we have defined  $\varphi_{k-1,1}: X_{k-1} \rightarrow Y_{k-1}$ , such that

$$\varphi_{k-1}(x_{k-1}, \varphi_{k-1,1}(x_{k-1})) \neq 0, \text{ where } x_{k-1} \in X_{k-1}.$$

Now define  $\varphi_{k1}: X_k \rightarrow Y_k$ ,

$$\text{if } x_k = f_k^n(x_n^0), \quad \varphi_{k1}(x_k) = g_k^n(y_n^0);$$

$$\text{if } x_k \neq f_k^n(x_n^0), \quad \varphi_{k1}(x_k) = y_{x_k}, \text{ where } y_{x_k} \in g_{k-1}^{-1}(\varphi_{k-1}(f_{k-1}(x_k)))$$

$$\text{such that } \varphi_k(x_k, y_{x_k}) \neq 0.$$

(such  $y_{x_k}$  exists, because

$$\varphi_k(x_k, g_{k-1}^{-1}(\varphi_{k-1,1}(f_{k-1}(x_k)))) = \varphi_{k-1}(f_{k-1}(x_k), \varphi_{k-1,1}(f_{k-1}(x_k))) \neq 0.)$$

So,  $\varphi_k(x_k, \varphi_{k1}(x_k)) \neq 0$ .

For each  $k = 1, \dots, n$ ,  $x_k \in X_k$ ,

$$g_{k-1} \circ \varphi_{k1}(x_k) = g_{k-1}(y_{x_k}) = \varphi_{k-1,1}(f_{k-1}(x_k))$$

$$\text{i.e. } g_{k-1} \circ \varphi_{k1} = \varphi_{k-1,1} \circ f_{k-1}.$$

$$\begin{array}{ccc} X_k & \xrightarrow{\varphi_{k1}} & Y_k \\ \downarrow f_{k-1} & & \downarrow g_{k-1} \\ X_{k-1} & \xrightarrow{\varphi_{k-1,1}} & Y_{k-1} \end{array}$$

$$\text{Define } \varphi_k^{(1)} = \frac{1}{1-t_1}(\varphi_k - t_1 \varphi_{k1}), \quad k = 0, 1, \dots, n.$$

For  $x_k \in X_k$ ,  $y_k \in Y_k$ , let  $x_n \in (f_k^n)^{-1}(x_k)$ , then

$$\varphi_k(x_k, y_k) = \varphi_n(x_n, (g_k^n)^{-1}(y_k)) \geq \varphi_n(x_n, y_n), \text{ any } y_n \in (g_k^n)^{-1}(y_k),$$

if  $\varphi_k(x_k, y_k) \neq 0$ , then we could pick  $y_n \in (g_k^n)^{-1}(y_k)$ , such that

$$\varphi_n(x_n, y_n) \neq 0, \text{ so } \varphi_k(x_k, y_k) \geq \varphi_n(x_n, y_n) \geq t_1;$$

if  $\varphi_k(x_k, y_k) = 0$ , then  $\varphi_{k1}(x_k) \neq y_k$ , i.e.  $\varphi_{k1}(x_k, y_k) = 0$ .

Therefore,  $\varphi_k^{(1)}(x_k, y_k) = \frac{1}{1-t_1} (\varphi_k(x_k, y_k) - t_1 \varphi_{k1}(x_k, y_k)) \geq 0$ , for

all  $x_k \in X_k, y_k \in Y_k$ .

$$\begin{aligned} \text{Also } \sum_{y_k \in Y_k} \varphi_k^{(1)}(x_k, y_k) &= \frac{1}{1-t_1} (\sum_{y_k \in Y_k} \varphi_k(x_k, y_k) - t_1 \sum_{y_k \in Y_k} \varphi_{k1}(x_k, y_k)) \\ &= \frac{1}{1-t_1} (1-t_1) = 1. \end{aligned}$$

Thus,  $\varphi_k^{(1)} \in \beta(m)(X_k, Y_k)$  and  $\varphi_k = t_1 \varphi_{k1} + (1-t_1) \varphi_k^{(1)}$ .

$g_k \circ \varphi_k^{(1)} = \varphi_{k-1}^{(1)} \circ f_k$ , since  $g_k \circ \varphi_k = \varphi_{k-1} \circ f_k$  and  $g_k \circ \varphi_{k1} = \varphi_{k-1,1} \circ f_k$ .

Repeat the same procedure to  $\{\varphi_k^{(1)}\}_{k=0, \dots, n}$ , if  $\varphi_n^{(1)}$  is not deterministic (from Lemma 2,  $\varphi_n^{(1)}$  is deterministic implies that  $\varphi_k^{(1)}$  is deterministic,  $k < n$ ), we could get  $\varphi_{k2}$  deterministic,  $s_2 \in (0, 1)$ , and  $\varphi_k^{(2)} \in \beta(m)(X_k, Y_k)$ , such that  $\varphi_k^{(1)} = s_2 \varphi_{k2} + (1-s_2) \varphi_k^{(2)}$ . Let  $t_2 = (1-t_1)s_2$ , then  $\varphi_k = t_1 \varphi_{k1} + t_2 \varphi_{k2} + (1-t_1-t_2) \varphi_k^{(2)}$ .

Then repeat the same procedure to  $\{\varphi_k^{(2)}\}, \dots$ . Since  $X_n, Y_n$  are finite, after finite steps, we get  $\varphi_k = \sum_{i=1}^s t_i \varphi_{ki}$ ,  $\varphi_{ki}$  are deterministic,  $i = 1, \dots, s, k = 0, \dots, n$ .

Lemma 4.  $Z = \{1, 2, \dots, s\}$  is a finite set,  $p$  is a measure on  $Z$  with  $p(Z) \leq 1$ . For  $0 < t < 1, 0 < \alpha \leq 1$ , and  $p(Z) > t\alpha$ , there exist measures  $q, q'$  on  $Z$ , with  $q(Z) = \alpha, q'(Z) = \frac{1}{1-t}(p(Z) - t\alpha)$ , such that  $p = tq + (1-t)q'$ .

Proof. 1). if  $\exists i \in Z$ , such that  $p(i) \geq t\alpha$ , then define

$$q(j) = \begin{cases} \alpha, & j = i \\ 0, & j \neq i \end{cases}$$

$$q'(j) = \begin{cases} \frac{1}{1-t}(p(i)-t\alpha), & j = i \\ \frac{1}{1-t} p(j), & j \neq i \end{cases}$$

since  $p(i) \geq t\alpha$ , so  $q'(i) \geq 0$ ,  $\sum_{j=1}^{\infty} q'(j) = \frac{1}{1-t}(p(i)-t\alpha) + \frac{1}{1-t} \sum_{j \neq i} p(j)$   
 $= \frac{1}{1-t}(p(Z)-t\alpha)$ ; and  $p = tq + (1-t)q'$ .

2). if  $\exists i \in Z$ , such that  $p(i) \geq (1-t)\alpha$ , then just interchange  $q$  and  $q'$  in case 1).

3). if for all  $i \in Z$ ,  $p(i) < \min\{t\alpha, (1-t)\alpha\}$ , since  $\frac{p(i)}{t} < \alpha$

and  $\sum_{i=1}^{\infty} \frac{p(i)}{t} = \frac{1}{t} \sum_{i=1}^{\infty} p(i) = \frac{1}{t} p(Z) > \alpha$ , so there exists  $k \in Z$ ,

such that  $\alpha - \sum_{i=1}^k \frac{p(i)}{t} \geq \frac{p(k)}{t}$  and  $\alpha - \sum_{i=1}^k \frac{p(i)}{t} < \frac{p(k+1)}{t}$ .

$$\text{Define } q(i) = \begin{cases} \frac{p(i)}{t}, & i \leq k \\ \alpha - \sum_{i=1}^k \frac{p(i)}{t}, & i = k+1 \\ 0, & i > k+1 \end{cases}$$

$$q'(i) = \begin{cases} 0, & i \leq k \\ \frac{1}{1-t}(p(k+1)-t(\alpha - \sum_{i=1}^k \frac{p(i)}{t})) = \frac{1}{1-t}(\sum_{i=1}^k p(i)-t\alpha), & i = k+1 \\ \frac{1}{1-t} p(i), & i > k+1 \end{cases}$$

Then,  $\sum_{i=1}^{\infty} q(i) = \sum_{i=1}^k \frac{p(i)}{t} + (\alpha - \sum_{i=1}^k \frac{p(i)}{t}) = \alpha$ ;

$$\begin{aligned} \sum_{i=1}^{\infty} q'(i) &= \frac{1}{1-t}(\sum_{i=1}^k p(i)-t\alpha) + \frac{1}{1-t} \sum_{i=k+1}^{\infty} p(i) = \frac{1}{1-t}(\sum_{i=1}^{\infty} p(i)-t\alpha) \\ &= \frac{1}{1-t}(p(Z) - t\alpha), \end{aligned}$$

and  $p = tq + (1-t)q'$ .

Lemma 5.  $X, Y, n$  and  $m$  are finite sets,  $f: X \rightarrow n, g: Y \rightarrow m$  are deterministic maps,  $\varphi: X \rightarrow Y, \psi: n \rightarrow m$  are random maps and the square

$$\begin{array}{ccc} X & \xrightarrow{\varphi} & Y \\ \downarrow & f & \downarrow g \\ n & \xrightarrow{\psi} & m \end{array} \quad \text{is commutative.}$$

If  $\psi = t\psi_1 + (1-t)\psi_2, 0 \leq t \leq 1$ , then there exist  $\varphi_1, \varphi_2 \in \mathcal{P}(m)(X, Y)$ , such that  $\varphi = t\varphi_1 + (1-t)\varphi_2$  and  $g \circ \varphi_i = \varphi_i \circ f, i = 1, 2$ .

Proof.  $g \circ \varphi = \psi \circ f$  and  $\psi = t\psi_1 + (1-t)\psi_2$  mean that for any  $x \in X, k \in m, \varphi(x, g^{-1}(k)) = \psi(f(x), k) = t\psi_1(f(x), k) + (1-t)\psi_2(f(x), k)$ . We wish to define  $\varphi_1, \varphi_2 \in \mathcal{P}(m)(X, Y)$ , such that

$$\varphi_i(x, g^{-1}(k)) = \psi_i(f(x), k), i = 1, 2, x \in X, k \in m.$$

If  $g^{-1}(k) = \emptyset$ , define  $\varphi_i(x, \emptyset) = 0$ , since  $\psi(f(x), k) = \varphi(x, g^{-1}(k)) = 0$ .

For  $g^{-1}(k) \neq \emptyset$ :

if  $\psi_2(f(x), k) = 0$ , then  $\varphi(x, g^{-1}(k)) = t\psi_1(f(x), k)$ ,

define  $\varphi_1(x, y) = \frac{\varphi(x, y)}{t}$  for  $y \in g^{-1}(k)$ , then  $\varphi_1(x, y) \geq 0$ ; since

$$\sum_{y \in g^{-1}(k)} \varphi_1(x, y) = \sum_{y \in g^{-1}(k)} \frac{\varphi(x, y)}{t} = \frac{1}{t} \varphi(x, g^{-1}(k)) = \frac{1}{t} (t\psi_1(f(x), k)) = \psi_1(f(x), k) \leq 1,$$

so  $\varphi_1(x, y) \leq 1$ ; define  $\varphi_2(x, y) = 0, y \in g^{-1}(k)$ ;

if  $\psi_2(f(x), k) \neq 0$ , take  $g^{-1}(k) = Z, p = \varphi(x, -), \alpha = \psi_2(f(x), k)$

then  $p(Z) > t\alpha$ . Using Lemma 4, we get  $\varphi_1(x, -), \varphi_2(x, -)$ , two measures on  $Z = g^{-1}(k)$ , and  $\varphi(x, y) = t\varphi_1(x, y) + (1-t)\varphi_2(x, y), y \in g^{-1}(k)$ .

Since  $Y$  is the disjoint union of  $g^{-1}(k), k \in m$  and measures are

finitely additive, we have the desired result.

Theorem. Let  $\varphi$  be a stochastic transformation from  $(X, f)$  to  $(Y, g)$  in  $S'$ . For any given  $n$ , we have decomposition  $\varphi = \sum_{i=1}^{m_n} t_i \varphi_{(i)}$ , such that  $\varphi_{(i)} \in S'((X, f), (Y, g))$  and  $(\varphi_{(i)})_k: X_k \longrightarrow Y_k$  is deterministic upto  $k \leq n$ .

Proof. For a given  $n$ , from Lemma 3, we have decomposition

$$\varphi_k = \sum_{i=1}^{m_n} t_i \varphi_{ki}, \quad k \leq n, \quad \varphi_{ki} \text{ are deterministic and}$$

$$E_{k-1} \varphi_{ki} = \varphi_{k-1, i} \circ f_{k-1}.$$

For  $k > n$ , repeatedly using Lemma 5, we get  $\varphi_k = \sum_{i=1}^{m_n} t_i \varphi_{ki}$ , and  $E_{k-1} \varphi_{ki} = \varphi_{k-1, i} \circ f_{k-1}$ .

Thus,  $\varphi_{(i)}$  with the  $k$ -th component  $(\varphi_{(i)})_k = \varphi_{ki}$  is a stochastic transformation from  $(X, f)$  to  $(Y, g)$ , i.e.  $\varphi_{(i)} \in S'((X, f), (Y, g))$ , and  $\varphi = \sum_{i=1}^{m_n} t_i \varphi_{(i)}$ .

Definition 2. We say that a stochastic transformation  $\varphi: (X, f) \longrightarrow (Y, g)$  is indecomposable, if there is no nontrivial decomposition of  $\varphi$ , i.e. for  $0 < t < 1$ ,  $\varphi = t\varphi_1 + (1-t)\varphi_2$  implies  $\varphi_1 = \varphi_2 = \varphi$ .

Examples.

Example 1.  $X_0, Y_0$  are two finite sets,  $\varphi_0: X_0 \longrightarrow Y_0$  is a random map. Define  $X_n = X_0^{n+1}$ ,  $Y_n = Y_0^{n+1}$  and  $\varphi_n = \otimes^{n+1} \varphi_0$ . We call  $\varphi$ , with the  $n$ th component  $\varphi_n$ , the stochastic transformation generated by  $\varphi_0$ .

If  $\varphi_0$  is not deterministic, from Lemma 1,  $\varphi_0 = \sum_{i=1}^m t_i \varphi^i$ ,  $\varphi^i$ ,  $i = 1, \dots, m$  are deterministic. Then  $\varphi_n = \otimes_{i=1}^{n+1} \varphi_0 = \otimes_{i=1}^{n+1} (\sum_{j=1}^m t_j \varphi^j)$   
 $= \sum_{k=1, \dots, m^{n+1}} t_{i_1} t_{i_2} \dots t_{i_{n+1}} \varphi^{i_1} \otimes \varphi^{i_2} \otimes \dots \otimes \varphi^{i_{n+1}}$ . We can not decompose  $\varphi$  into finite many stochastic transformations  $\varphi_{(i)}$ , such that  $\varphi_{(i)}$  are indecomposable.

Example 2.  $\varphi : X \rightarrow Y$  is in  $S^n$ , i.e.  $X_n = X_0^{n+1}$ ,  $Y_n = Y_0^{n+1}$ ,  $X_0, Y_0$  are finite sets. If for some  $n$ , as  $k > n$ ,  $\varphi_{k+1} = \varphi_k \otimes \varphi'$ , where  $\varphi' : X_0 \rightarrow Y_0$  is a deterministic map, using Lemma 3, we get  $\varphi_k = \sum_{i=1}^m t_i \varphi_{ki}$ ,  $k \leq n$ ,  $\varphi_{ki}$  deterministic.

Then  $\varphi_{n+p} = \varphi_n \otimes (\otimes^p \varphi') = \sum t_i \varphi_{ni} \otimes (\otimes^p \varphi')$ , denote  $\varphi_{ni} \otimes (\otimes^p \varphi') = \varphi_{n+p,i}$ ,  $\varphi_{n+p,i}$  is deterministic, since  $\varphi'$  and  $\varphi_{ni}$  are deterministic.

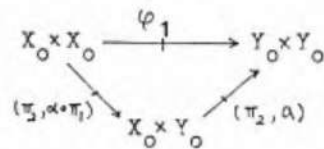
Therefore,  $\varphi_{(i)}$ , with the  $k$ -th component  $(\varphi_{(i)})_k = \varphi_{ki}$ , is a stochastic transformation from  $X$  to  $Y$ ,  $i = 1, \dots, m$ , and  $\varphi = \sum_{i=1}^m t_i \varphi_{(i)}$ ,  $\varphi_{(i)}$  is indecomposable.

Example 3.  $X_0, Y_0$  are finite sets,  $\alpha : X_0 \rightarrow Y_0$ ,  
 $a : X_0 \times Y_0 \rightarrow Y_0$  are random maps.

Define  $X_n = X_0^{n+1}$ ,  $Y_n = Y_0^{n+1}$  ;

$\varphi_0 = \alpha : X_0 \rightarrow Y_0$  ,

$\varphi_1 = (\pi_2, a) \circ (\pi_2, \alpha \circ \pi_1) : X_0 \times X_0 \rightarrow Y_0 \times Y_0$



assume  $\varphi_{n-1}: X_0^n \rightarrow Y_0^n$  is defined,

then define  $\varphi_n = (\pi_{2\dots n+1}, a \circ \pi_{12}) (\pi_{n+1}, \varphi_{n-1} \circ \pi_{(n)})$

$$\begin{array}{ccc}
 X_0^{n+1} & \xrightarrow{\varphi_n} & Y_0^{n+1} \\
 \searrow (\pi_{n+1}, \varphi_{n-1} \circ \pi_{(n)}) & & \nearrow (\pi_{2\dots n+1}, a \circ \pi_{12}) \\
 X_0^n \times Y_0^n & & 
 \end{array}$$

$$\begin{aligned}
 \text{Then } \pi_{(n)} \circ \varphi_n &= \pi_{(n)} \circ (\pi_{2\dots n+1}, a \circ \pi_{12}) \circ (\pi_{n+1}, \varphi_{n-1} \circ \pi_{(n)}) \\
 &= \varphi_{n-1} \circ \pi_{(n)}
 \end{aligned}$$

i.e. the following square is commutative.

$$\begin{array}{ccc}
 X_0^{n+1} & \xrightarrow{\varphi_n} & Y_0^{n+1} \\
 \pi_{(n)} \downarrow & & \downarrow \pi_{(n)} \\
 X_0^n & \xrightarrow{\varphi_{n-1}} & Y_0^n
 \end{array}$$

$\varphi_{n-1}$  has the property, for  $(x_1, \dots, x_n) \in X_0^n$ ,  $(y_1, \dots, y_n) \in Y_0^n$ ,  
 $\varphi_{n-1}(x_1, \dots, x_n; y_1, \dots, y_n) = \alpha(x_1, y_1) a(x_2, y_1, y_2) a(x_3, y_2, y_3) \dots a(x_n, y_{n-1}, y_n)$ .

For any  $n$ , using Lemma 3, we get  $\varphi_k = \sum_{i=1}^{m_n} t_i \varphi_{ki}$ ,  $k \leq n$ , where  
 $\varphi_{ki}$  are deterministic and  $\pi_{(k)} \circ \varphi_{ki} = \varphi_{k-1, i} \circ \pi_{(k)}$ .

Decompose  $a$  by  $a = \sum_{j=1}^s s_j a_j$ , where  $a_j$  are deterministic.

$$\begin{aligned}
 \text{Then } \varphi_{n+1} &= (\pi_{2\dots n+2}, a \circ \pi_{12}) \circ (\pi_{n+2}, \varphi_n \circ \pi_{(n+1)}) \\
 &= (\pi_{2\dots n+2}, (\sum_{j=1}^s s_j a_j) \circ \pi_{12}) \circ (\pi_{n+2}, (\sum_{i=1}^{m_n} t_i \varphi_{ni}) \circ \pi_{(n+1)}) \\
 &= \sum_{j,i} s_j t_i (\pi_{2\dots n+2}, a_j \circ \pi_{12}) (\pi_{n+2}, \varphi_{ni} \circ \pi_{(n+1)}) .
 \end{aligned}$$

If  $a$  is not deterministic, we run into the same difficulty

as in Example 1.

If  $a$  is deterministic, then we can decompose  $\varphi$  into finite many indecomposable stochastic transformations.

$a$  is deterministic, then  $(\pi_{2\dots n+p}, a \circ \pi_{12})$  is deterministic.

$$\text{Then } \varphi_{n+1} = \sum_{i=1}^{m_n} t_i (\pi_{2\dots n+2}, a \circ \pi_{12}) \circ (\pi_{n+2}, \varphi_{ni} \circ \pi_{(n+1)})$$

denote  $(\pi_{2\dots n+2}, a \circ \pi_{12}) \circ (\pi_{n+2}, \varphi_{ni} \circ \pi_{(n+1)}) = \varphi_{n+1,i}$ ,  $\varphi_{n+1,i}$  are deterministic and  $\varphi_{n+1} = \sum_{i=1}^{m_n} t_i \varphi_{n+1,i}$ .

Assume  $\varphi_{n+p} = \sum_{i=1}^{m_n} t_i \varphi_{n+p,i}$ ,  $\varphi_{n+p,i}$  are deterministic.

$$\begin{aligned} \text{then } \varphi_{n+p+1} &= (\pi_{2\dots n+p+2}, a \circ \pi_{12}) \circ (\pi_{n+p+2}, \varphi_{n+p} \circ \pi_{(n+p+1)}) \\ &= (\pi_{2\dots n+p+2}, a \circ \pi_{12}) \circ (\pi_{n+p+2}, (\sum_{i=1}^{m_n} t_i \varphi_{n+p,i}) \circ \pi_{(n+p+1)}) \\ &= \sum_{i=1}^{m_n} t_i (\pi_{2\dots n+p+2}, a \circ \pi_{12}) \circ (\pi_{n+p+2}, \varphi_{n+p,i} \circ \pi_{(n+p+1)}) \end{aligned}$$

so  $\varphi_{n+p+1,i} = (\pi_{2\dots n+p+2}, a \circ \pi_{12}) \circ (\pi_{n+p+2}, \varphi_{n+p,i} \circ \pi_{(n+p+1)})$  are deterministic and  $\varphi_{n+p+1} = \sum_{i=1}^{m_n} t_i \varphi_{n+p+1,i}$ .

It is easy to check that for  $k > n$ ,  $i = 1, \dots, m_n$ , we also have

$\pi_{(k)} \circ \varphi_{ki} = \varphi_{k-1,i} \circ \pi_{(k)}$ . Thus,  $\varphi_{(i)}$ , with  $k$ -th component  $(\varphi_{(i)})_k = \varphi_{ki}$  is a stochastic transformation from  $X$  to  $Y$ , and  $\varphi = \sum_{i=1}^{m_n} t_i \varphi_{(i)}$ .

#### §4. Stochastic dynamic programming.

The essential characteristic of stochastic dynamic programming is that the dynamics (or transformation) not just depend on the states but also on decisions. It has been called sequential decision problem or controlled stochastic process.

##### Formulation (discrete).

$$\mathbb{N} = 0, 1, \dots$$

1). Spaces of states, at each stage  $n$  the space of states is  $X_n$ ,  $n \in \mathbb{N}$ .

2). Spaces of decisions, at each stage  $n$  the space of available decisions is  $A_n$ ,  $n \in \mathbb{N}$ .

3). Dynamic (transformation, law of the motion), at each stage  $n$  is a map  $T_n: \prod_{k \in \mathbb{N}} X_k \times \prod_{k \in \mathbb{N}} A_k \longrightarrow X_{n+1}$ ;

in Markov case, dynamic does not depend on the history

$$T_n: X_n \times A_n \longrightarrow X_{n+1}.$$

4). Policy (decision rule) is a law of choosing decisions, at each stage  $n$ ,  $\delta_{n-1}: \prod_{k \in \mathbb{N}} X_k \times \prod_{k \in \mathbb{N}} A_k \longrightarrow A_n$ .

Remark. a). In [2], Bellman's definition, he used stationary Markov case, i.e.  $X_n = X$ ,  $A_n = A$  for all  $n \in \mathbb{N}$ , and

$T_n: X \times A \longrightarrow X$ ,  $\delta_n: X \times A \longrightarrow A$ . Same is in Gihman[6], state space  $X_n = (X, \mathcal{A})$ , decision space  $A_n = (U, \mathcal{B})$ , for all  $n \in \mathbb{N}$ .

b). If we require  $T_n, \delta_n$  in  $\mathcal{M}$  (the category of measurable

spaces and measurable functions), we get deterministic dynamic programming (see [2]); if we require  $T_n, \delta_n$  in  $\mathcal{P}(M)$  (the category of measurable spaces and random maps), then we get stochastic dynamic programming, this is the case in [3] and [6].

5). Criteria of optimal policy (best decision rule).

$g: X_k \times A_k \rightarrow K$  is the reward function at stage  $k$ , where  $K$  is the nonnegative reals. Whenever it is appropriate to consider that the reward of process is the sum of the rewards of its stages then it is also appropriate to define the following functional of  $T$ ,  $f_N(x_0) = \max_{a_0, a_1, \dots, a_{N-1}} \sum_{k=0}^{N-1} g(x_k, a_k)$ , where  $x_k = T(x_{k-1}, a_{k-1})$ .

If a policy is given, i.e. on each stage  $n$ ,  $\delta_{n-1}: \prod_{k=0}^{n-1} X_k \times \prod_{k=0}^{n-1} A_k \rightarrow A_n$  is given, and an initial state  $x_0 \in X_0$  is chosen, then we have a sequence  $x_0, a_0 = \delta_0(x_0), x_1 = T_0(x_0, a_0), a_1 = \delta_1(x_0, x_1, a_0), \dots, x_n = T_{n-1}(x_0, \dots, x_{n-1}, a_0, \dots, a_{n-1}), a_n = \delta_{n-1}(x_0, \dots, x_n, a_0, \dots, a_{n-1}), \dots$ .  $\delta^*$  is an optimal policy if for any  $N \in \mathbb{N}$ , and such a sequence corresponding to  $\delta^*$ , we have  $\sum_{k=0}^{N-1} g(x_k, a_k) = f_N(x_0)$ .

In stochastic case:  $T_n: \prod_{k=0}^n X_k \times \prod_{k=0}^n A_k \rightarrow X_{n+1}$  is a random map, for each  $(x_0, \dots, x_n, a_0, \dots, a_n)$  in  $\prod_{k=0}^n X_k \times \prod_{k=0}^n A_k$ , there is an induced probability measure  $T_n(\cdot | x_0, \dots, x_n, a_0, \dots, a_n)$  on  $X_{n+1}$  and for  $\delta_n: \prod_{k=0}^n X_k \times \prod_{k=0}^n A_k \rightarrow A_n$ , there is

an induced probability measure  $\delta_n(\cdot | x_0, \dots, x_n, a_0, \dots, a_{n-1})$  on  $A_n$ . So  $T_n, \delta_n$  and an initial probability measure  $\mu$  on  $X_0$  give a probability measure  $\Delta_{\mu, \delta}$  on  $X \times A = \prod_{k=0}^{\infty} X_k \times \prod_{k=0}^{\infty} A_k$ .

Let  $(\prod_{k=0}^n B_k \times C_k) \times (\prod_{k=0}^n X_k \times A_k) = B_0 \times C_0 \times \dots \times B_n \times C_n \times X_{n+1} \times A_{n+1} \dots$  be a measurable cylinder in  $X \times A$ , where  $B_k, C_k$  are measurable sets in  $X_k, A_k$  respectively,  $k = 0, \dots, n$ , then

$$\begin{aligned} & \Delta_{\mu, \delta}((\prod_{k=0}^n B_k \times C_k) \times (\prod_{k=0}^n X_k \times A_k)) \\ = & \int_{B_0} \int_{C_0} \dots \int_{B_n} \int_{C_n} \delta_n(da_n | x_0, \dots, x_n, a_0, \dots, a_{n-1}) T_{n-1}(dx_n | x_0, \dots, x_{n-1}, \\ & a_0, \dots, a_{n-1}) \dots \delta_1(da_0 | x_0) \mu(dx_0). \end{aligned}$$

If  $\mu$  is a probability measure on  $\prod_{k=0}^{\infty} X_k \times \prod_{k=0}^{\infty} A_k$ , then  $\Delta_{\mu, \delta}^P$  is a probability measure on  $X \times A$ , such that

$$\begin{aligned} & \Delta_{\mu, \delta}^P((\prod_{k=0}^n B_k \times C_k) \times (\prod_{k=0}^n X_k \times A_k)) \\ = & \int_{\prod_{k=0}^n B_k \times \prod_{k=0}^n C_k} \int_{C_p} \dots \int_{B_n} \int_{C_n} \delta_n(da_n | x_0, \dots, x_n, a_0, \dots, a_{n-1}) T_{n-1}(dx_n | x_0, \dots, \\ & x_{n-1}, a_0, \dots, a_{n-1}) \dots \delta_p(da_p | x_0, \dots, x_p, a_0, \dots, a_{p-1}) d\mu \end{aligned}$$

At each stage  $n$ , there is a loss function (or cost, reward, return)  $c_n: X_n \times A_n \rightarrow K$ , it is a measurable function.

Then  $c(x, a) = \sum_{n=0}^{\infty} c_n(x_n, a_n)$ , where  $(x, a) \in X \times A$ ,

$$c_p^n(x, a) = \sum_{k=p}^n c_k(x_k, a_k)$$

the risk 
$$L_p^n(\mu, \delta) = E_{\Delta, \mu, \delta}^p (c_p^n(x, a)) = \int_{X \times A} c_p^n(x, a) d\Delta_{\mu, \delta}^p .$$

Let  $\mu(x_0, \dots, x_n, a_0, \dots, a_{n-1})$  be the probability measure on

$\prod_{k=0}^n X_k \times \prod_{k=0}^{n-1} A_k$ , which is concentrated on  $(x_0, \dots, x_n, a_0, \dots, a_{n-1})$ .

Then a policy  $\delta^*$  is said to be optimal at stage  $n$ , if

$$L_n(\mu(x_0, \dots, x_n, a_0, \dots, a_{n-1}), \delta^*) = \inf_{\delta} L_n(\mu(x_0, \dots, x_n, a_0, \dots, a_{n-1}), \delta)$$

(for reward or return functions we should consider

$$L_n(\mu(x_0, \dots, x_n, a_0, \dots, a_{n-1}), \delta^*) = \sup_{\delta} L_n(\mu(x_0, \dots, x_n, a_0, \dots, a_{n-1}), \delta))$$

for all  $(x_0, \dots, x_n, a_0, \dots, a_{n-1}) \in \prod_{k=0}^n X_k \times \prod_{k=0}^{n-1} A_k$ .

A policy is fully optimal, if it is optimal at any stage, see [3] , [6] .

The existence of a (fully) optimal policy depends on a). under certain topology, the space of policies is compact; b). the risk function is lower semicontinuous with respect to this topology, then use the theorem: a lower semicontinuous real valued function defined on a compact space assumes its inf.

#### Discussion.

1). About cost functions.

A cost function is trying to tell us some kind relation or measurement about what was happening and the decision we made. This external measurement should reflect, reveal the more deep internal relation, the causation. Objectively there is a direct connection between state space  $X$  and decision space  $A$ ,  $\hat{f}: X \rightarrow A$ .

We could obtain it through observation, experiments, analysing and synthesizing various external measurements.

Because of the complexity of the real world, noise observation and possible mistakes,  $l$  also could be random.

2). For convenience, we would sometimes like policy also to be defined on  $\prod_{k \in n} X_k \times \prod_{k \in n} A_k$ . Once the dynamic  $T_n$  is given, and a policy  $\delta_n$  is chosen at each stage  $n$ , we can form the composition of following two random maps, and define it as the policy at stage  $n$ .

$$\prod_{k \in n} X_k \times \prod_{k \in n} A_k \xrightarrow{(1_{\prod_{k \in n} X_k}, T_n, 1_{\prod_{k \in n} A_k})} \prod_{k \in n} X_k \times X_{n+1} \times \prod_{k \in n} A_k \xrightarrow{\delta_n} A_{n+1}$$

3). Policy depends on partial informations.

Since our observation is partial and our decisions are made just based on our observation, so the decision rule should just depend on this partial information, but not the whole state space. In [6], p.36, "Control of processes with incomplete observation", state space  $X$  with  $\sigma$ -algebra  $\mathcal{B}$ , decision space  $A$  with  $\sigma$ -algebra  $\mathcal{A}$ ;  $\delta_n$ , the policy at stage  $n$ , is defined on  $Y_n \times A^{n-1}$ , where  $Y_n = (X^n, \mathcal{B}')$ ,  $\mathcal{B}'$  is a sub  $\sigma$ -algebra of  $\mathcal{B}^n$ , i.e.  $\delta_n$  only depends on part of the state space  $(X^n, \mathcal{B}^n)$ . Note that, this relation between state space  $(X^n, \mathcal{B}^n)$  and observation (partial information)  $Y_n = (X^n, \mathcal{B}')$  is just a measurable function  $1_{X^n}: (X^n, \mathcal{B}^n) \longrightarrow (X^n, \mathcal{B}')$ .

Thus, generally we use a measurable function  $f_n: \prod_{k \in n} X_k \longrightarrow Y_n$  to describe the relation of state space  $\prod_{k \in n} X_k$  and the space of observable states  $Y_n$  at each stage  $n$ .  $f_n$  also could be random,

which means that the observation is obtained through a noisy channel. Note that,  $Y_n$  may not be a product, even if the states are. Thus, the policy at stage  $n$ , is a random map  $\delta_n: Y_n \times \prod_{k=0}^n A_k \rightarrow A_{n+1}$ .

4). More generally, also for simplicity, we could take  $X_n$  and  $A_n$  as subspaces of  $\prod_{k=0}^n X_k$  and  $\prod_{k=0}^n A_k$ , at each stage  $n$ , such that  $X_n$ 's and  $A_n$ 's are compatible with projections.

After these discussions, we give another formulation of stochastic dynamic programming.

1). state space  $X_n$  for each  $n$ ,  $X_n$  includes the history until stage  $n$ .

2). decision space  $A_n$  for each  $n$ ,  $A_n$  includes the history until stage  $n$ .

3). the partial information obtained from  $X_n$ ,  $f_n: X_n \rightarrow Y_n$ .

4). dynamic (law of the motion) at each stage  $n$ ,

$$q_n: X_n \times A_n \rightarrow X_{n+1}$$

5). policy at stage  $n$ ,  $\delta_n: Y_n \times A_n \rightarrow A_{n+1}$ .

6). the recall of history,

$$h_n: X_{n+1} \rightarrow X_n, \quad \varepsilon_n: A_{n+1} \rightarrow A_n,$$

$h_n, \varepsilon_n$  are deterministic.

Remark. In the previous case,  $\pi_{(n)}: \prod_{k=0}^n X_k \rightarrow \prod_{k=0}^n X_k$  is the recall of the history, and  $\pi_{(n)} \circ T_n = 1_{\prod_{k=0}^n X_k}$ .

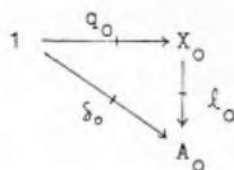
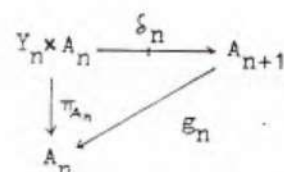
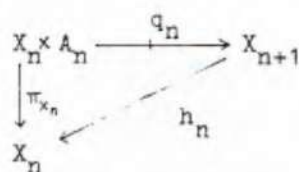
7). initial probability measure on  $X_0$ ,  $q_0: 1 \rightarrow X_0$ ;

initial probability measure on  $A_0$ ,  $\delta_0: 1 \rightarrow A_0$ .

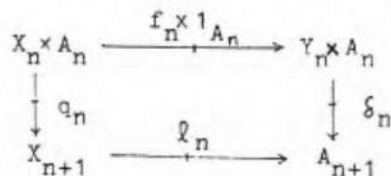
8). the correct action at each stage  $n$ ,

$$\ell_n: X_n \rightarrow A_n.$$

The following diagrams are commutative.



9). the optimal policy. In the following square



we want  $\text{dist}(\delta_n \circ f_n \times 1_{A_n}, \ell_n \circ q_n)$  as small as possible, i.e.  $\delta^*$  is an optimal policy, if

$$\text{dist}(\delta_n^* \circ f_n \times 1_{A_n}, \ell_n \circ q_n) = \inf_{\delta} \text{dist}(\delta_n \circ f_n \times 1_{A_n}, \ell_n \circ q_n)$$

for each  $n$ .

Existence of such an optimal policy.

Lemma. Let  $X, Y, Z$  be measurable spaces and  $h: Z \rightarrow Y$  is deterministic. Let  $\Delta_h = \{g \in \mathcal{P}(m)(X \times Y, Z) \mid h \circ g = \pi_Y\}$  i.e.  $g \in \Delta_h$ ,

if  $X \times Y \xrightarrow{\varepsilon} Z$  is commutative.

$$\begin{array}{ccc} X \times Y & \xrightarrow{\varepsilon} & Z \\ \downarrow \pi_Y & \swarrow h & \\ Y & & \end{array}$$

Then  $\Delta_h$  is a closed subset of  $\mathcal{P}(m)(X \times Y, Z)$  with contravariant topology induced by the convex metric.

Proof. Let  $\delta \in \bar{\Delta}_h$ , the closure of  $\Delta_h$  in  $\mathcal{P}(m)(X \times Y, Z)$ , for any  $\varepsilon > 0$ ,  $g \in B(\varepsilon, \delta) \cap \Delta_h$ , i.e.  $\exists t > e^{-\varepsilon}$ ,  $g = t\delta + (1-t)\delta'$ , then  $\pi_Y = h \circ g = t(h \circ \delta) + (1-t)(h \circ \delta')$ .

For any measurable set  $B$  in  $Y$ ,  $\pi_Y((x, y), B) = \begin{cases} 1, & y \in B \\ 0, & y \notin B \end{cases}$ ,

this implies  $h \circ \delta((x, y), B) = \begin{cases} 1, & y \in B \\ 0, & y \notin B \end{cases}$ , i.e.  $\pi_Y = h \circ \delta$ , so  $\delta \in \Delta_h$ .

From section 1, we know that the function

$$\text{dist}(\cdot, f_n \times 1_{A_n}, l_n \circ q_n): \mathcal{P}(m)(Y_n \times A_n, A_{n+1}) \longrightarrow K$$

is continuous respect to the topology induced by the convex metric, so restrict it to  $\Delta_{\varepsilon_n} = \{\delta_n \in \mathcal{P}(m)(Y_n \times A_n, A_{n+1}) \mid \varepsilon_n \circ \delta_n = \pi_{A_n}\}$ , it

is still continuous.

If  $\mathcal{P}(m)(Y_n \times A_n, A_{n+1})$  is compact, then so is  $\Delta_{\varepsilon_n}$ , as in case  $Y_n, A_n, A_{n+1}$  are finite sets; If  $\mathcal{P}(m)(Y_n \times A_n, A_{n+1})$  is not compact, under some conditions the subspace  $\Delta_{\varepsilon_n}$  may be compact, or we just work with a compact subset of  $\Delta_{\varepsilon_n}$ , for each  $n$ . Then there is an optimal policy.

#### Reference

1. Robert B. Ash, Real analysis and probability, New York, Academic Press, 1972.
2. Bellman and Kalaba, Dynamic programming and modern control theory, Academic Press, 1965.
3. Lawrence D. Brown and Bharat T. Doshi, Existence of optimal policy in stochastic dynamic programming, Probability and Mathematical Statistics, Vol.1, Fasc.2(1980). p.171-184.
4. N.N.Cencov, Statistical decision rules and optimal inference, Translations of Mathematical Monographs, Volume 53, American Mathematical Society, 1982.
5. S.Eilenberg and G.M.Kelly, Closed category, Proceeding of Conference on Categorical Algebra, La Jolla, 1965.
6. I.I.Gihman and A.V.Skorohod, Controlled stochastic processes Springer-Verlag, 1979.
7. Alex Heller, Probabilistic automata and stochastic transformations, Mathematical System Theory, Vol.1, No. 3.p.197-280, 1967.
8. J.R.Isbell, M.I.Klun and S.H.Schanuel, Affine parts of algebraic theories.I, Journal of Algebra, Vol.44,No.1,January 1977.
9. J.R.Isbell, M.I.Klun and S.H.Schanuel, Affine parts of algebraic theories. II, Can. J. Math., Vol.XXX, No.2,1978, p.231-237.
10. F.W.Lawvere, The category of probabilistic mapping, 1962.

11. F.W.Lawvere, Metric spaces, generalized logic, and closed categories, dell'Universita di Perugia, 1973.
12. James L.Melsa and David L.Cohn, Decision and estimation theory, McGaw-Hill Book Company, 1978.
13. Ioana Schiopu, A categorical approach to probability theory, Thesis,
14. Zbigniew Semadeni, Banach spaces of continuous functions, Volume 1, PWN-Polish Scientific publishers, Warazawa 1971.
15. D.J.White, Dynamic programming, Oliver & Boyd, 1969.
16. Eugene M. Zaustinsky, Space with non-symmetric distance, Memoirs of the American Mathematical Society, #34, 1959.