



HAL
open science

On Stochastic Transformation Models.

Han-Ping Li

► **To cite this version:**

| Han-Ping Li. On Stochastic Transformation Models.. 1997. hal-00129566

HAL Id: hal-00129566

<https://hal.science/hal-00129566>

Preprint submitted on 8 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Stochastic Transformation Models

Han-Ping LI

Université Louis Pasteur, Strasbourg, France

Résumé

In this paper, the nonparametric transformation model is introduced and studied in comparison with the more familiar parametric transformation model and semiparametric transformation model. The semiparametric transformation model and nonparametric transformation model are characterized in term of stochastic transformations, with the independence in the first model, without the independence in the second one. The semiparametric transformation model is also characterized as the weak cloture of the convex envelope of the parametric transformation model. The LeCam deficiency is also considered.

Key Words: Group Model, Transformation Model, Deficiency, Comparison of Experiments.

AMT Codes: 62B10, 62E10

Résumé

In this paper, the nonparametric transformation model is introduced and studied in comparison with the more familiar parametric transformation model and semiparametric transformation model. The semiparametric transformation model and nonparametric transformation model are characterized in term of random transformations, with the independence in the first model, without the independence in the second one. The semiparametric transformation model is also characterized as the weak cloture of the convex envelope of the parametric transformation model. The LeCam deficiency is also considered.

AMS classification : 62B10, 62E10

Key words : Group Model, Transformation Model, Deficiency, Comparison of Experiments.

1 Introduction

Many statistical problems exhibit some symmetries, which provide natural restrictions to impose invariance properties under a suitable group of transformations on statistical procedures to be used. Closed related to this concept is the notion of parametric group model. See for example, [1], [2], and [3].

First of all, we shall state the basic assumptions used throughout the paper and introduce some notations.

Let (Ω, \mathcal{A}) be a measurable space and let G be a group acting on it endowed with its Borel field $\sigma(G)$ with respect to some topology. We shall assume that

(1) G is a LCD space (c.f. [3].), that is, G is a locally compact group with a denumerable basis for the topology.

(2) The map $(g, \omega) \mapsto g\omega$ is $\sigma(G) \otimes \mathcal{A}$ measurable.

(3) There is a measurable equivariant S from (Ω, \mathcal{A}) to $(G, \sigma(G))$ such that

$$S(g\omega) = gS(\omega).$$

Remarks :

1) If $\mathcal{L}(X) = P$, then $\mathcal{L}(gX) = Pg^{-1}$.

2) Equivariant estimators are often used in statistics. Suppose that the model $(\Omega, \mathcal{A}, \mathcal{P})$ is closed under the group actions, that is, $\forall g \in G$

$$Pg^{-1} \in \mathcal{P} \text{ whenever } P \in \mathcal{P}.$$

Consider $\delta(X)$ as an estimator of a functional $T(P)$. Naturally $T(Pg^{-1})$ could be estimated by $\delta(gX)$ as well as by $g\delta(X)$. Equating these two quantities, we should have

$$\delta(gx) = g\delta(x).$$

3) Equivariants do exist if the action of G on Ω is free, that is, for all $x \in \Omega$, if $y \in Gx$ then there exists only one $g \in G$ such that $y = gx$. Under this condition, we can choose an orbit representation: for each orbit Gx of Ω , we select one point $T(x)$ to represent that orbit Gx . $T(x)$ is obviously a maximal invariant. Since $x \in GT(x)$ and

G acts freely, then there exists a unique element $S(x) \in G$ such that $x = S(x)T(x)$. It is easy to verify that $S(x)$ is an equivariant.

Under these assumptions, we can define a maximal invariant T from (Ω, \mathcal{A}) to itself by

$$T(\omega) \triangleq S(\omega)^{-1}\omega.$$

It is easy to verify that T is a measurable maximal invariant, that is,

$$\begin{aligned} T(g\omega) = T(\omega) \text{ and } T(\omega) = T(\omega') \\ \text{implies } g \circ \omega = \omega' \text{ for some } g \in G. \end{aligned}$$

A maximal invariant is able to help us to describe the orbits under the group actions, that is, for every subset B of Ω , we have

$$\{\omega \in \Omega \mid T(\omega) \in T(B)\} = \bigcup_{g \in G} g^{-1}(B).$$

In fact, if $\omega \in g^{-1}(B)$ for some $g \in G$, then $g\omega \in B$, so $T(\omega) = T(g\omega) \in T(B)$.

On the other hand, if $T(\omega) \in T(B)$, then there is a $\omega' \in B$ such that $T(\omega) = T(\omega')$. Since T is a maximal invariant, there is a $g \in G$ with $g\omega = \omega' \in B$, so $\omega \in g^{-1}B$.

Denote by $\text{Cap}(Y, \mathcal{B})$, the Class of All Probability measures defined on (Y, \mathcal{B}) .

Let P_0 be a probability measure defined on (Ω, \mathcal{A}) . The group G generates in a natural way the following well known model:

(1) **Parametric transformation model:**

$$\begin{aligned} \mathcal{M}_p(P_0) \triangleq \{P_g \in \text{Cap}(\Omega, \mathcal{A}) : P_g(B) = P_0\{g^{-1}(B)\}, \\ \forall B \in \mathcal{A}, g \in G\} \end{aligned}$$

For a family of measures $\mathcal{P} = \{P\}$, we can just concentrate our attention on each orbit: $\mathcal{M}_p(P)$, $P \in \mathcal{P}$.

The parametric transformation model can be mixed by a priori law from the Bayesian point of view, that is, this model in his turn generates also the following model:

(2) **Semiparametric transformation model :**

$$\mathcal{M}_s(P_0) \hat{=} \{P_\mu : P_\mu(B) = \int P_g(B)\mu(dg), \\ \forall B \in \mathcal{A}, \mu \in \text{Cap}(G, \sigma(G))\}$$

Inspired by the idea of Rüschemdorf (c.f. [5]) : , we introduce in this paper :

(3) **Nonparametric transformation model :**

$$\mathcal{M}_n(P_0) \hat{=} \{P \in \text{Cap}(\Omega, \mathcal{A}) : PT^{-1} = P_0T^{-1}\}$$

Roughly speaking, this is the maximal family in which the T remains ancillary. It is clear that for all $P_g \in \mathcal{M}_p(P)$, we have $P_gT^{-1} = P_0T^{-1}$ with T defined as above (1.2). Same holds for $P_\mu \in \mathcal{M}_s(P)$. Therefore, we have

$$\mathcal{M}_p(P_0) \subset \mathcal{M}_s(P_0) \subset \mathcal{M}_n(P_0)$$

We shall study in this paper some relationships and the characterization problem of the three models. Before doing this, let us just consider some examples.

Example 1 (One dimensional location-scale model)

Let $\mathbf{1}$ denote the vector of 1's in \mathbf{R}^n , $\Omega = \mathbf{R}^n \setminus \text{span}\{\mathbf{1}\}$, and let P be a probability measure defined on $(\Omega, \mathcal{B}(\Omega))$, and dominated by the Lebesgue measure μ , with $\frac{dP}{d\mu}(x) = f(x)$.

Let $G = \{g = (\sigma, \mu), \sigma \in R^+, \mu \in R\}$ with action $g(x) = (\sigma x_1 + \mu, \dots, \sigma x_n + \mu)$, for $x = (x_1, \dots, x_n) \in \Omega$; and the group operation being

$$(\sigma_2, \mu_2) \circ (\sigma_1, \mu_1) = (\sigma_2\sigma_1, \sigma_2\mu_1 + \mu_2).$$

The orbits of all group action for a partition of Ω , and

$$S(x) = (s, \bar{x})$$

is an equivariant, with

$$\bar{x} \hat{=} \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad s \hat{=} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

$$T(x) = S(x)^{-1}x = \left(\frac{x_1 - \bar{x}}{s}, \dots, \frac{x_n - \bar{x}}{s} \right)$$

is the maximal invariant. It is easy to see that

$$\mathcal{M}_p(P) = \{P_{\sigma, \mu}, \sigma \in \mathbf{R}^+, \mu \in \mathbf{R}\}$$

with $\frac{dP_{\sigma, \mu}}{d\mu}(x) = f((\sigma, \mu)^{-1}x) = f\left(\frac{x_1 - \mu}{\sigma}, \dots, \frac{x_n - \mu}{\sigma}\right).$

Example 2 (General location-scale model)

Let $(H, \langle \cdot, \cdot \rangle)$ be an Hilbert space and $L \subset H$ be a proper linear subspace of H . Denote P_L the orthogonal projection operator on L , Id the identity application. Let $G = \mathbf{R}^+ \times L$ operate on $\Omega = H \setminus L$ by the operation

$$[\sigma, h]x = \sigma x + h, \quad h \in L, \sigma > 0,$$

then $S(x) = [s(x), P_L(x)]$ is an equivariant, where $s(x) = \sqrt{\frac{1}{n-1} \langle x - P_L(x), x - P_L(x) \rangle}$. It is clear that $T = S^{-1}Id = [s(x)^{-1}, -s(x)^{-1}\bar{x}(x)]$ is the maximal invariant :

$$T(x_1, \dots, x_n) = S(x)^{-1}x = \left(\frac{x_1 - P_L(x)}{s(x)}, \dots, \frac{x_n - P_L(x)}{s(x)} \right)$$

In particular, if $H = \mathbf{R}^n$ and $L = \text{span}\{\mathbf{1}\}$, we reind the result of the example 1.

Example 3 (Group model)

$\Omega = G$, $S = Id$ is an equivariant,
and , $T = e$ is the maximal invariant.

2 Main results and proofs

Suppose now that the space Ω is separable. Denote by $Z \sim P$ to mean that the law of the variable Z is P , that is,

$$P(B) = Pr(\omega \in \Omega : Z(\omega) \in B).$$

Denote T the maximal invariant defined as before, then we have:

Theorem 1 $P \in \mathcal{M}_n(P_0)$ iff there are X, Y defined on a probability sapce $(\tilde{\Omega}, \mathcal{B}, \mathbf{P})$ such that

$$\mathcal{L}(X | \mathbf{P}) = P_0, \quad \mathcal{L}(Y | \mathbf{P}) = P, \quad \mathbf{P}(T(X) = T(Y)) = 1.$$

Proof.

”Only if part : ” We give two different proofs here.

(1) Let $\Lambda = \{(\omega, \omega') \in \Omega \times \Omega \mid \exists g \in G, \omega' = g \circ \omega\}$. A consequence of Strassen’s theorem is the following :

$$\mathbf{P}(\Lambda) = 1 \iff \sup \left\{ P(U) - P_0 \left(\pi_1(\Lambda \cap (\Omega \times U)) \right) \mid U \subset \Omega \text{ is open} \right\} = 0$$

with $\pi_1 : \Omega \times \Omega \rightarrow \Omega$ the first projection. Since

$$\begin{aligned} & \pi_1 \{ \Lambda \cap (\Omega \times U) \} \\ &= \pi_1 \{ (\omega, \omega') \in \Omega \times \Omega \mid \exists g \in G, \omega' = g \circ \omega \in U \} \\ &= \{ \omega \in g^{-1}(U) \text{ for some } g \in G \} \\ &= \cup_{g \in G} g^{-1}(U) \\ &= \{ \omega \in \Omega \mid T(\omega) \in T(U) \} \quad \text{by (1.2)}. \end{aligned}$$

Therefore,

$$\begin{aligned} & \sup \left\{ P(U) - P_0 \left(\pi_1(\Lambda \cap (\Omega \times U)) \right) \right\} \\ &= \sup \left\{ P(U) - P_0 \left(T^{-1}(T(U)) \right) \right\} \\ &\leq \sup \left\{ P \left(T^{-1}(T(U)) \right) - P_0 \left(T^{-1}(T(U)) \right) \right\} = 0. \end{aligned}$$

(2) Suppose $P \in \mathcal{M}_n(P_0)$. Define a set function \hat{P} on the set of rectangular sets by :

$$\hat{P}(A \times B) = \int P_0(A|T = s)P(B|T = s)PT^{-1}(ds) \\ \forall A, B \in \mathcal{A}$$

It is clear that \hat{P} can be extended on $(\mathcal{A} \times \mathcal{A})$ as a probability measure with :

$$\mathbf{P}(A \times B|T = s) = P_0(A|T = s)P(B|T = s).$$

Let $X : \Omega \times \Omega \rightarrow \Omega$, $Y : \Omega \times \Omega \rightarrow \Omega$ be the first projection and the second projection respectively, then we have immediately :

$$\mathcal{L}(X) = P_0, \text{ and } \mathcal{L}(Y) = P$$

due to $P_0T^{-1} = PT^{-1}$.

We have furthermore :

$$\begin{aligned} \mathbf{P}(A \times B) &= \int P_0(A|T = s)P(B|T = s)PT^{-1}(ds) \\ &= \int P(1_B(s)P_0(A|T = s)|T = s)PT^{-1}(ds) \\ &= \int 1_B(s)P_0(A|T = s)PT^{-1}(ds) \\ &= \hat{P}(T^{-1}(A) \cap T^{-1}(B)) \\ &= \hat{P}(T(X) \in A \cap T(Y) \in B). \end{aligned}$$

The second and third equalities being true because $P_0T^{-1} = PT^{-1}$. Since the space is separable, we then have $\mathbf{P}(T(X) \neq T(Y)) = 0$.

” If part : ”

$$\begin{aligned} &PT^{-1}(B) \\ &= P(T(Y) \in B) = \mathbf{P}(T(U(X)) \in B) \\ &= \mathbf{P}(T(X) \in B) \\ &= \mathbf{P}(X \in T^{-1}(B)) = \hat{P}T^{-1}(B). \end{aligned}$$

Based on this theorem, we have

Theorem 2 (*Stochastic Transformation Characterization*)

(1) $P \in \mathcal{M}_s(P_0)$ iff there are $X \sim P_0, U : \Omega \rightarrow G$ which is independent of X , such that $Y = U \circ X \sim P$.

(2) $P \in \mathcal{M}_n(P_0)$ iff there are $X \sim P_0, U : \Omega \rightarrow G$, such that $Y = U \circ X \sim P$.

Proof:

(1) "Only if part: "

Suppose $P \in \mathcal{M}_s(P_0)$, then there is a probability measure $\mu \in \text{Cap}(G, (\mathcal{G}))$ such that $P(B) = \int P_g(B)\mu(dg)$. Define a probability measure

$\tilde{P} \triangleq P \otimes \mu$ on $(\Omega \times G, \mathcal{A} \otimes \sigma(G))$, and maps $X : \Omega \times G \rightarrow \Omega$, $U : \Omega \times G \rightarrow G$ the projections, then we have immediately that $X \sim P, U \sim \mu$ which is independent of X , and for $Y = U(X)$,

$$\begin{aligned} \tilde{P}(Y \in B) &= \tilde{P}(U(X) \in B) \\ &= \int_G \tilde{P}(U(X) \in B | U = g) \tilde{P}U^{-1}(dg) \\ &= \int_G \tilde{P}(X \in g^{-1}(B) | U = g) \mu(dg) \\ &= \int_G P(X \in g^{-1}(B)) \mu(dg) \\ &= \int_G P_g(B) \mu(dg) = P(B). \end{aligned}$$

The forth equality being true because X and GU are independent.

" If part: "

$$\begin{aligned} P(B) &= \tilde{P}(U(X) \in B) \\ &= \int \tilde{P}(U(X) \in B | U = g) \tilde{P}U^{-1}(dg) \\ &= \int \tilde{P}(X \in g^{-1}(B)) \tilde{P}U^{-1}(dg) \\ &= \int P_g(B) \tilde{P}U^{-1}(dg) \in \mathcal{M}_s(P). \end{aligned}$$

(2) "If part: "

If there are $X \sim P, U : \Omega \rightarrow G$, such that $Y = U \circ X \sim Q$, then $T(Y) = T(U(X)) = T(X)$ since T is an maximal invariant for G . Application of the first theorem gives the conclusion.

” Only if part : ”

If $S(X)^{-1} \circ X = T(X) = T(Y) = S(Y)^{-1} \circ Y$, then $Y = S(Y) \circ S(X)^{-1} \circ X = U \circ (X)$.

Theorem 3 (*Weak closure Characterization*)

$$\mathcal{M}_s(P) = \overline{\text{Conv}}(\mathcal{M}_p(P))$$

where Conv denotes the convex envelope, and the closure is taken w.r.t. weak topology on $\text{Cap}(\Omega, \mathcal{A})$.

Proof:

If $P \in \mathcal{M}_s(P_0)$, then

$$\begin{aligned} Q(B) &= R(U(X) \in B) \\ &= \int R(U(X) \in B | U = g) R U^{-1}(dg) \\ &= \int R(X \in g^{-1}(B)) R U^{-1}(dg) = P * \mu(B) \end{aligned}$$

with $\mu = R U^{-1}$. let $E \subset G$ be a countable dense set of G , the by the theorem 6.3 , p 44 of [6], there is a sequence of measures whose supports are finite subsets of E , converge weakly to μ . that is,

$$\mu_n = \sum_{i=n_1}^{n_k} \alpha_i g_i \implies \mu$$

Consequently, $Q = P * \mu = \text{Lim} P * \mu_n \in \overline{\text{Conv}}(\mathcal{M}_p(P))$.

If $Q \in \overline{\text{Conv}}(\mathcal{M}_p(P))$, then there is a sequence

$$Q_n = \sum_{i=n_1}^{n_k} \alpha_i P_{g_i} \in \text{Conv}(\mathcal{M}_p(P))$$

such that Q_n converges weakly to Q . Q_n can be also written as

$$Q_n = \int P_g \mu_n(dg) = P * \mu_n$$

with $\mu_n = \sum_{i=1}^{n_k} \alpha_i g_i \in \text{Conv}(\theta, \theta \in G)$. It follows from the theorem 2.2 p 59 of [4], that $\{\mu_n\}$ is relatively compact. Denote μ the limit point of $\{\mu_n\}$, we have

$$Q(B) = \int P_g(B) \mu(dg)$$

by theorem 1.1 ,p 57 of [4] .

LeCam's deficiency is a famous statistical distance. Let " \supset^ϵ " denote the relation of " ϵ -informative" c.f. [7] and [8]. Then

Theorem 4 (ϵ -deficiency)

$$\mathcal{M}_p(P) \supset^\epsilon \mathcal{M}_p(Q) \quad \text{iff} \quad \mathcal{M}_s(P) \supset^\epsilon \mathcal{M}_s(Q)$$

Proof:

We use the invariant kernal criterion (c.f. Torgensen [8]), that is:

$$\mathcal{M}_p(P) \supset^\epsilon \mathcal{M}_p(Q)$$

if and only if there exists a invariant kernal K such that , $\|KP - Q\| \leq \epsilon$, where K being invariant means that $K(x, B) = K(gx, gB) \quad \forall g \in G$.

Since $KP_g(B) = \int K(x, B)P_g(dx) = \int K(gx, B)P(dx) = \int K(x, g^{-1}B)P(dx)$, we have

$$\|KP_g - Q_g\| = \sup_B |KP_g(B) - Q_g(B)| \leq \epsilon \text{ for all } g \in G.$$

Therefore

$$\begin{aligned} \|KP_\mu - Q_\mu\| &= \left\| \int (KP_g - Q_g) \mu(dg) \right\| \\ &\leq \int \|KP_g - Q_g\| \mu(dg) \leq \epsilon. \end{aligned}$$

Theorem 5 (ϵ -deficiency)

$$\mathcal{M}_p(P) \supset^\epsilon \mathcal{M}_p(Q) \quad \text{iff} \quad \mathcal{M}_n(P) \supset^\epsilon \mathcal{M}_n(Q).$$

3 BIBLIOGRAPHY

- [1] Chang U., Eaves D. (1990) *Reference prior for the orbit in a group model. Ann. Statist.* **18**, No 4, 1595-1614.
- [2] Eaton, Moris L.(1989) *Group Invariance Applications in Statistics. Regional Conference Serightarrowes in Probability and Statistics* , Vol.1, 145 pp.
- [3] Jespersen, Niels Ch. B.(1989) *On the structure of transformation models. Ann. Statist.* **17**, 195-208.
- [4] Parthasarathy, K.R.(1967) *Probability Measures on Metric Spaces. Academic Press.*
- [5] Rûschendorf, L.(1989) *On random translation models. Statist. & Probab. Letters North-Holland* **7** 361-367.
- [6] Strassen,V. (1965) *The existence of probability measures with given marginals. Ann. Math. Statist.***36**, 423-439
- [7] Strasser,H. (1985) *Mathematical Theory of Statistics. De Gruyter, Berlin.*
- [8] Torgensen,E.N. (1972) *Comparison of translation experiments. Ann. Math. Statist.* **43**, 1383-1391

Han-Ping LI

Département de Mathématique

Université Louis Pasteur

7 rue René-Descartes

67084 Strasbourg France