Characterization of statistical models for which a given statistic has the same distribution.

Han-Ping Li

Characterization of statistical models for which a given statistic has the same distribution

Han-Ping LI
Université Louis Pasteur, Strasbourg, France
Université de Wuhan, Wuhan, Chine
E-mail: lihp@irma.u-strasbg.fr

Abstract

In this paper, a rather general method to characterize the statistical models for which a given statistic has the same distribution is given. First, the semi-parametric transformation model and non-parametric transformation model are characterized in term of random transformations, with the independence in the first model, without the independence in the second one. By applying these characterizations, the problem of describing all all distributions for which the distribution of T is the same can be solved by first finding a suitable group of transformations for which T is a maximal invariant and then applying a random transformation on a suitable random variable. Many statistical applications are given, including the characterization of the statistical models for which a Fisher's or Student's statistic has the same distribution.
AMS classification : 62E10, 62H05, 62A25
Key words : Group Model, Transformation Model, Distributions Characterization, Random Transformation

1 Introduction

In the areas of estimation and hypothesis testing problems, the optimal solutions are easily found if the underlying population has normal distributions. In this case,

the distributions of some important statistics for inferential purposes, such as classical t- and F- statistics are available. B. Efron (1969) pointed out that the weaker condition such as some symmetry can successfully replace the normal condition as far as the classical t-statistics is concerned.

This is not an exceptional case. Many statistical problems exhibit some symmetries with respect to a group of transformations. These symmetries not only provide natural restrictions to impose invariance properties on statistical procedures to be used, but also give some simple ways to characterize the underlying statistical models.

More precisely, the group of transformations can generate in different ways some statistical models which are closed under any transformation of the group. Given a sample space, given a group of transformation actions on the left of it, a non-parametric transformation model is defined and studied in comparison with the more familiar parametric transformation model and semi-parametric one. Some of their relationships and characterizations are established here.

As applications, we can easily give the characterization of the statistical models for which a given statistic has the same distribution. For a given statistics T, we try to find a suitable group of transformations for which T is a maximal invariant. Then the characterization theorem about non-parametric transformation model generated by this maximal invariant permit us to find the looking class via some random transformations.

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

Let $(\Omega, \mathcal{A})$ be a general measurable space and let G be a group acting on the left of Ω. We shall assume that

(1) $(\Omega, \mathcal{A})$ is a Hausdorff locally compact measurable space.

(2) G is a Hausdorff locally compact group with a countable topological base (LCD space), endowed with its Borel field $\sigma(G)$.

(3) The map $(g, \omega) \longmapsto g \circ \omega \hat{=} g\omega$ is $\sigma(G) \otimes \mathcal{A}$ measurable.

In most applications, we shall assume in addition that

(4^*) There exists almost surely a maximal invariant T from $(\Omega, \mathcal{A})$ to a measurable space $(\Omega, \mathcal{A})$ such that there exists a measurable subset Ω_0 with $\mathbf{P}(\Omega_0) = 1$ and that

$$T(g\omega) = T(\omega) \quad \text{for } g \in G, \omega \in \Omega_0 \text{ and } T(\omega) = T(\omega') \text{ imply } g\omega = \omega'\, \omega, \omega' \in \Omega_0, \text{ for some } g \in G.$$

Denote by $\mathcal{L}(Z \mid \mathbf{P}) = \mathbf{Q}$ to mean that the law of the variable Z is P, that is,

$$\mathbf{Q}(B) = \mathbf{P}(\omega \in \Omega : Z(\omega) \in B).$$

Remark:

1. If $\mathcal{L}(X \mid \mathbf{P}) = P$ then $\mathcal{L}(gX \mid \mathbf{P}) = Pg^{-1}$ where Pg^{-1} denote the distribution P lifted by g: $Pg^{-1}(B) = P\big(g^{-1}\{B\}\big)$, for every $B \in \mathcal{A}$.

2. In practice, we often use maximal invariants introduced by an equivariant, that is, a measurable S from $(\Omega, \mathcal{A})$ to $\big(G, \sigma(G)\big)$ such that

$$\Big(S(g\omega)\Big)^{-1} \circ (g\omega) = \Big(S(\omega)\Big)^{-1} \circ \omega \text{ a.s.}$$

   From S, a maximal invariant T can be introduced by by

$$T(\omega) \hat{=} \Big(S(\omega)\Big)^{-1} \circ \omega.$$

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

3. Maximal invariants and equivariants exist if the action of G on Ω is free, that is, for all $x \in \Omega$, if $y \in Gx = \{g\,x, g \in G\}$ 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 this 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.

4. 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

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

The notion of a maximal invariant enables us to describe the orbits under the group actions. In general, if T is a invariant, then for every subset B of Ω,

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

This can be showed as this: if $\omega \in g^{-1}(B)$ for some $g \in G$, then $g\omega \in B$, so $T(\omega) = T(g\omega) \in T(B)$.

If in addition, T is a maximal invariant, then we have

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

In fact, 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 $\mathrm{Cap}(Y, \mathcal{B})$, the class of all probability measures defined on $(Y, \mathcal{B})$.

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

(1) **Parametric transformation model :**

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

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

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) **Semi-parametric transformation model :**

$$\begin{aligned}\mathcal{M}_s(P_0) \quad \hat{=}&\{P_\mu : P_\mu(B) = \int P_g(B)\mu(\mathrm{d}\, g),\\ &\forall B \in \mathcal{A}, \mu \in \mathrm{Cap}(G, \sigma(G))\}\end{aligned}$$

Inspired by the idea of Rûschendorf (c.f. [7, 1989].), we introduce in this paper:

(3) **Non-parametric transformation model :**

$$\begin{aligned}\mathcal{M}_n(P_0) \quad \hat{=}&\{P \in \mathrm{Cap}(\Omega, \mathcal{A}) : PT^{-1} = P_0T^{-1}\}\\ &\text{where } T \text{ is a maximal invariant.}\end{aligned}$$

Roughly speaking, this is the maximal family in which the distribution of T remains the same. (So T is ancillary). It is this property that permit us to solve the characterization problems.

In many cases, we shall inverse the path: given a statistics T, we try to find a group of transformations for which T is a maximal invariant. Then the characterization theorem about $\mathcal{M}_n(P_0)$ permit us to find the class of statistical models admitting the same distribution for T.

It is clear that for all $P_g \in \mathcal{M}_p(P)$, we have $P_gT^{-1} = P_0T^{-1}$. The 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 these three models. Before doing this, let us just consider some examples.

Example 1 (n-dimensional location-scale model)

Let $\mathbf{1}$ denote the vector of 1's in $\Omega = \mathbf{R}^n$, $\Omega = \mathbf{R}^n$. Let $G = \{g = (\sigma, \mu), \quad \sigma \in R^+, \mu \in R\}$ with action $g(\mathbf{x}) = \sigma\mathbf{x}+\mu\mathbf{1} = (\sigma x_1+\mu, ..., \sigma x_n+\mu)$, for $\mathbf{x} = (x_1, ..., x_n) \in \Omega$; and the group operation being

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

Denote $\Omega_0 = \mathbf{R}^n \setminus \text{span}\{\mathbf{1}\}$. Since the span$\{\mathbf{1}\}$ is closed for G, so $g\omega \in \Omega_0$ whenever $\omega \in \Omega_0$. let P be a probability measure defined on $(\Omega, \mathcal{B}(\Omega))$ with $P(\Omega_0) = 1$ and dominated by the Lebesgue measure μ, with $\frac{dP}{d\mu}(\mathbf{x}) = f(\mathbf{x}), \mathbf{x} = (x_1, ..., x_n) \in \Omega$.

The orbits of all the group actions form a partition of Ω_0 and

$$S(\mathbf{x}) = (s(\mathbf{x}), \overline{x}(\mathbf{x}))$$

is an equivariant, with

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

well defined on Ω_0.

It is clear that the inverse of S($\mathbf{x}$) on Ω_0 is equal to $S(\mathbf{x})^{-1} = (s^{-1}(\mathbf{x}), -s^{-1}(\mathbf{x})\overline{x})$, and that

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

is a maximal invariant. It is easy to see that

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

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

Example 2 (General location-scale model)

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

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

The group operation is

$$[\sigma_2, h_2] \circ [\sigma_1, h_1] = [\sigma_2\sigma_1, \sigma_2 h_1 + h_2].$$

let P be a probability measure defined on $(\Omega, \mathcal{B}(\Omega))$ such that $P(\Omega_0) = 1$ with $\Omega_0 = H \setminus L$, then $S(x) = [s(x), P_L(x)]$ is an equivariant well defined on Ω_0, where

$s(x) = \sqrt{\frac{1}{n-1} < x - P_L(x), x - P_L(x) >}$. It is clear that the inverse of S(**x**) is equal to $S(\mathbf{x})^{-1} = (s(x)^{-1}, -s(x)^{-1}P_L(x))$, and that

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

is the maximal invariant.

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

Example 3 (Group model)

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

$\mathbf{P}T^{-1} = \delta_e$, therefore, $M_n(P_0) = \text{Cap}(G, \sigma(G))$.

Example 4 (Rank statistics)

$\Omega = \mathbf{R}^n, G = \{g : \mathbf{R} \longmapsto \mathbf{R} \text{ is strictly increasing . } g \circ (x_1, \cdots, x_n) = (g(x_1), \cdots g(x_n))$.

Equivariant doesn't exist, but if we define $r_i = 1 + \sum_{k=1}^{n} I_{\{x_k < x_i\}}$, then $T(x_1, \cdots, x_n) = (r_1, \cdots, r_n)$ is a maximal invariant.

Example 5 (n-dimensional normal scale model)

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

$$\sigma_2 \circ \sigma_1 = \sigma_2\sigma_1.$$

Denote $\Omega_0 = \mathbf{R}^n \setminus \{\mathbf{0}\}$. let $P_0 = \mathcal{N}(0,1)^{\otimes n}$ be the normal probability measure defined on $(\Omega, \mathcal{B}(\Omega))$, we have $P_0(\Omega_0) = 1$.

For $\mathbf{x} = (x_1, ..., x_n) \in \Omega$, denote $f(\mathbf{x})$, the density $\frac{\mathrm{d}P_0}{\mathrm{d}\mu}(\mathbf{x})$ with respect to the Lebesgue measure μ .

The orbits of all the group actions form a partition of Ω_0 and

$$S(\mathbf{x}) = \|\mathbf{x}\|$$

is an equivariant, with

$$\|\mathbf{x}\| \hat{=} \sqrt{\sum_{i=1}^{n} x_i^2}.$$

It is clear that on Ω_0 the inverse of S(**x**) is equal to $S(\mathbf{x})^{-1} = \|\mathbf{x}\|^{-1}$, and that

$$T(\mathbf{x}) = S(\mathbf{x})^{-1}\mathbf{x} = \left(\frac{x_1}{\|\mathbf{x}\|}, ..., \frac{x_n}{\|\mathbf{x}\|}\right)$$

is a maximal invariant. It is easy to see that

$$\mathcal{M}_p(P_0) = \{P_\sigma,\ \sigma \in\ R^+\ \} = \{\mathcal{N}(0,\sigma^2)^{\otimes n}\ \sigma \in\ R^+\}.$$
$$\text{with} \quad \frac{\mathrm{d}\, P_\sigma}{\mathrm{d}\,\mu}(\mathbf{x}) = f((\sigma)^{-1}\mathbf{x}) = f\left(\frac{x_1}{\sigma}, ..., \frac{x_n}{\sigma}\right).$$

Remark:

According to the theorem 2, if $\Gamma P = P$ for all orthogonal transformations Γ, then P belongs to $\mathcal{M}_n(P_0)$.

Since the Student's t statistics is a function of T so the distribution of T determines the distribution of the Student's t statistics. therefore the Student's t statistics still has the Student's distribution.

2 Some theorems and proofs

Suppose now that the space Ω is separable metric space. Suppose that all probability measures P under discussion are tight, that is, for all $\epsilon > 0$, there is a measurable compact set K_ϵ such that $P(K_\epsilon) > 1 - \epsilon$. This will be the case if in addition Ω is complete separable metric space.

Denote T an invariant with respect to the group G, then we have:

Theorem 1 *Suppose T is an invariant. If there exist two random variables X, Y defined on a probability space $(\tilde{\Omega}, \mathcal{B}, \mathbf{P})$ such that*

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

then $P \in \mathcal{M}_n(P_0)$.

Proof.

$$\begin{aligned} PT^{-1}(B) &= \mathbf{P}(T(Y) \in B) = \mathbf{P}(T(X) \in B) \\ &= \mathbf{P}(X \in T^{-1}(B)) = P_0 T^{-1}(B). \end{aligned}$$

Theorem 2 *Suppose T is a maximal invariant. Then $P \in \mathcal{M}_n(P_0)$ iff there exist X, Y defined on a probability space $(\tilde{\Omega}, \mathcal{B}, \mathbf{P})$ such that*

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

Proof.
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 [8, 1965]. is the following :

$$\mathbf{P}(\Lambda) = 1 \Longleftrightarrow$$
$$\sup\Big\{ P(U) - P_0\Big(\pi_1\big(\Lambda \cap (\Omega \times U)\big)\Big) \; U \subset \Omega \;\text{ is open } \Big\} = 0$$

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

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

therefore,

$$\begin{aligned} & \sup\Big\{ P(U) - P_0\Big(\pi_1\big(\Lambda \cap (\Omega \times U)\big)\Big) \\ = \;& \sup\Big\{ P(U) - P_0\Big(T^{-1}\big(T(U)\big)\Big) \Big\} \\ & \le \sup\Big\{ P\Big(T^{-1}\big(T(U)\big)\Big) - P_0\Big(T^{-1}\big(T(U)\big)\Big)\Big) \Big\} = 0. \end{aligned}$$

(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 to $(\mathcal{A} \times \mathcal{A})$ as a probability measure with :

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

Let $X : \Omega \times \Omega \mapsto \Omega$, $Y : \Omega \times \Omega \mapsto \Omega$ 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 $\mathbf{P}(T(X) \neq T(Y)) = 0$.

It is clear that if $P \in \mathcal{M}_p(P_0)$ iff there exist X with $\mathcal{L}(X \mid \mathbf{P}) = P_0$ and a transformation $g : \Omega \to G$ such that if we define $Y = g \circ X$, then $\mathcal{L}(Y \mid \mathbf{P}) = P$. Due to the previous theorem, we have

Theorem 3 *(Stochastic Transformation Characterization)*
Suppose T is an invariant. Then (1) $P \in \mathcal{M}_s(P_0)$ if there exist X with $\mathcal{L}(X \mid \mathbf{P}) = P_0$ and a stochastic transformation $U : \Omega \to G$ independent of X, such that

$\mathcal{L}(Y \mid \mathbf{P}) = P$ *where* $Y = U \circ X$.
(2) $P \in \mathcal{M}_n(P_0))$ *if there exist* X *with* $\mathcal{L}(X \mid \mathbf{P}) = P_0$ *and a stochastic transformation* $U : \Omega \to G$, *such that*$\mathcal{L}(Y \mid \mathbf{P}) = P$ *where* $Y = U \circ X$.

Proof :
(1)

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

(2)
If there exist $\mathcal{L}(X \mid \mathbf{P}) = P$ and $U : \Omega \to G$, such that $\mathcal{L}(Y \mid \mathbf{P}) = Q$ for $Y = U \circ X$, then $T(Y) = T(U(X)) = T(X)$ since T is an invariant for G. Applying the previous theorem gives the conclusion.

Theorem 4 *(Stochastic Transformation Characterization)*
Suppose T *is a maximal invariant. Then (1)* $P \in \mathcal{M}_s(P_0)$ *iff there exist* X *with* $\mathcal{L}(X \mid \mathbf{P}) = P_0$ *and a stochastic transformation* $U : \Omega \to G$ *independent of* X, *such that* $\mathcal{L}(Y \mid \mathbf{P}) = P$ *where* $Y = U \circ X$.
(2) $P \in \mathcal{M}_n(P_0))$ *iff there exist* X *with* $\mathcal{L}(X \mid \mathbf{P}) = P_0$ *and a stochastic transformation* $U : \Omega \to G$, *such that* $\mathcal{L}(Y \mid \mathbf{P}) = P$ *where* $Y = U \circ X$.

Proof :
(1) "Only if part : "
Suppose $P \in \mathcal{M}_s(P_0)$, then there exists a probability measure $\mu \in \mathrm{Cap}(G, (\mathcal{G}))$ such that $P(B) = \int P_g(B)\mu(\mathrm{d}\, g)$. Define a probability measure $\tilde{P} \hat{=} P \otimes \mu$ on $(\Omega \times G, \mathcal{A} \otimes \sigma(G))$, and the projections $X : \Omega \times G \mapsto \Omega$ and $U : \Omega \times G \mapsto G$, then we have immediately that $\mathcal{L}(X \mid \mathbf{P}) = P_0$, $\mathcal{L}(U \mid \mathbf{P}) = \mu$ is independent of X, and for

$Y = U(X)$,

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

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

(2) "Only if part : "

By the precedent theorem, there is a measurable set Ω_0 with $P(\Omega_0) = 1$ such that $T(X) = T(Y)$ holds on Ω_0. If $\omega \in \Omega_0$, we have $T\big(X(\omega)\big) = T\big(Y(\omega)\big)$. Since T is a maximal invariant, there is a $g(\omega)$ such that $Y(\omega) = g(\omega) \circ X(\omega)$. Define $U : \omega \longmapsto g(\omega)$ on Ω_0, then $Y = U \circ X$ on Ω_0.

It follows from this theorem that the problem of describing all all distributions for which the distribution of T is the same can be solved by first finding a suitable group of transformations for which T is a maximal invariant and then applying a random transformation on a suitable random variable.

3 Some applications

It follows from Theorem 2 that the problem of describing all all distributions for which the distribution of T is the same can be solved by first finding a suitable group of transformations for which T is a maximal invariant and then applying a random transformation on a suitable random variable.

In this section, we shall inverse the path: given a statistics T, we try to find a suitable group of transformations for which T is a maximal invariant. Then the characterization theorem about $\mathcal{M}_n(P_0)$ permit us to find all distributions admitting the same distribution for T.

Recall that a random vector X is said to have a spherically symmetric distribution if for every orthogonal transformation Γ, we have $\mathcal{L}(\Gamma X) = \mathcal{L}(U)$. It is well

known that X has a spherically symmetric distribution if and only if X has a spherical representation, that is, $\mathcal{L}(X) = \mathcal{L}(RU)$ where U has the uniform distribution on on the unit sphere $S_{n-1} \subset \mathbf{R}^n$, and $R \geq 0$ is a non negative random variable independent of U. The variable $U \in S_{n-1}$ is said ti have the uniform distribution on the unit sphere S_{n-1} if and only if $\mathcal{L}(\Gamma U) = \mathcal{L}(U)$ for every orthogonal transformation Γ. The simplest random vector having spherically symmetric distribution is just normal vector with $\mathcal{N}(0, I_n)$ distribution. In this case, the density of R with respect to Lebesgue measure is given by:

$$f(r) = \begin{cases} \frac{r^{n-1}}{2^{n-1}\Gamma(n/2)} \exp(-r^2/2) & \text{if } r > 0. \\ 0 & \text{otherwise} \end{cases}$$

Example 6
Let $\Omega = \mathbf{R}^n$, $G = \{g = \sigma : \ \sigma \in R^+\}$ with action $g(\mathbf{x}) = \sigma\mathbf{x} = (\sigma x_1, ..., \sigma x_n)$, for $\mathbf{x} = (x_1, ..., x_n) \in \Omega$; and the group operation being

$$\sigma_2 \circ \sigma_1 = \sigma_2\sigma_1.$$

let $P_0 = \mathcal{N}(0, 1)^{\otimes n}$ be the normal probability measure defined on $(\Omega, \mathcal{B}(\Omega))$, we have $P_0(\Omega_0) = 1$ with $\Omega_0 = \mathbf{R}^n \setminus \{\mathbf{0}\}$. It is clear that $T(\mathbf{x}) = \dfrac{\mathbf{x}}{\| \mathbf{x} \|}$ is a maximal invariant. Note that if $\mathcal{L}(\mathbf{x}) = P_0$, and if T' is a random transformation, then $T'\mathbf{x} = T'RU$ can again be represented as $R'U$ with $R' = T'R$ and U two random variables not necessarily independent. Therefore,

$$\mathcal{M}_n(P_0) = \{\mathcal{L}(RU), \ R \geq 0, U \ \text{ has uniform distribution on } S_{n-1}\}.$$

We can easily construct random transformations to get all spherically symmetric distributions from $\mathcal{N}(0, I_n)$ distribution. To see this, note that if $\mathcal{L}(X) = \mathcal{N}(0, I_n)$ then the density function of R in the spherical representation is given by

$$f(r) = \begin{cases} \frac{r^{n-1}}{2^{n-1}\Gamma(n/2)} \exp(-r^2/2) & \text{if } r > 0. \\ 0 & \text{otherwise} \end{cases}$$

which gives a strictly increasing the cumulated distribution function $F_1(t)$ on $(0, \infty)$.Denote this distribution by $P_2 = \mathcal{L}(R)$. Now let Y be a random variable having a spherically symmetric distribution with the spherical representation $\mathcal{L}(X) = \mathcal{L}(RU)$. Denote the cumulated distribution function of R by $F_2(t)$, define Define $\Omega = S_{n-1} \times R^+$ and

$\mathbf{P} = \mathcal{U}(S_{n-1}) \otimes P_2$. and $X : (s,r) \longmapsto rs$ and $U : (s,r) \longmapsto \frac{F_2(r)}{F_1(r)}$. It is clear that $\mathcal{L}(X) = \mathcal{N}(0,1)$, and the distribution of $U \circ X$ is exactly that of Y. That show directly that all spherically symmetric distributions belong to $\mathcal{M}_n(P_0)$.

Note that the Student's statistic $\frac{\sqrt{n}\overline{x}}{s_x}$ is a function of the maximal invariant $T(\mathbf{x}) = \dfrac{\mathbf{x}}{\| \mathbf{x} \|}$, hence all the distributions in $\mathcal{M}_n(P_0)$ which includes all spherically symmetric distributions will give the same distribution for the Student's t statistic $\frac{\sqrt{n}\overline{x}}{s_x}$.

Example 7

Let $\Omega = \mathbf{R}^n$, $P_0 = \mathcal{N}(0,1)^{\otimes n}$ be the normal probability measure defined on $(\Omega, \mathcal{B}(\Omega))$, we have $P_0(\Omega_0) = 1$ with $\Omega_0 = \mathbf{R}^n \setminus \{\mathbf{0}\}$. Consider the t- statistic $T(\mathbf{x}) = \dfrac{\langle \mathbf{1}, \mathbf{x} \rangle}{\| \mathbf{x} \|} = \dfrac{\sum_{i=1}^n x_i}{\| \mathbf{x} \|}$ where $\mathbf{1} = (1, \cdots, 1)$, and $\mathbf{x} = (x_1, \cdots, x_n)$. We look for a suitable group for which T will be a maximal invariant. It can be verified that G is generated by two subgroups G_1 and G_2 with $G_1 = \{g = \sigma : \quad \sigma \in R^+\}$, and $G_2 = \{g = \Gamma : \quad \Gamma \in O_n, \text{with } \Gamma\mathbf{1} = \mathbf{1}\}$. By applying the theorem, we have

$$\mathcal{M}_n(P_0) = \{\mathcal{L}(R\Gamma U), \ R \geq 0, \Gamma\mathbf{1} = \mathbf{1}, U \ \text{ has uniform distribution on } S_{n-1}\}.$$

This is exactly the characterization of all distributions for which the Student's statistic has the same distribution.

Example 8

Let $\Omega = \mathbf{R}^n$, $P_0 = \mathcal{N}(0,1)^{\otimes n}$ be the normal probability measure defined on $(\Omega, \mathcal{B}(\Omega))$, we have $P_0(\Omega_0) = 1$ with $\Omega_0 = \mathbf{R}^n \setminus \{\mathbf{0}\}$. Consider the F-statistic

$F(\mathbf{x}) = \dfrac{\| A\mathbf{x} \| / a}{\| B\mathbf{x} \| / b}$ where A and B are two projection matrices with $AB = 0$, and rank a and b respectively. Since $AB = 0$, we have $a + b \leq n$. So $\mathbf{R}^n = F_1 \oplus F_2 \oplus F_3$ with $F_1 = Im(A)$, $F_2 = Im(B)$, and $F_3 = Im((A \oplus B)^{\perp})$. Here again, we look for a suitable group for which F will be a maximal invariant.

Suppose for $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{y} \in \mathbf{R}^n$, we have $\| A\mathbf{x} \| = \| A\mathbf{y} \|$ and $\| B\mathbf{x} \| = \| B\mathbf{y} \|$. Write down the orthogonal decomposition of $\mathbf{x}$ and $\mathbf{y}$ with $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3$ and $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 + \mathbf{y}_3$, we have

$\| \mathbf{x}_1 \| = \| A\mathbf{x} \| = \| \mathbf{y}_1 \|$, and $\| \mathbf{x}_2 \| = \| B\mathbf{x} \| = \| \mathbf{y}_2 \|$. Since $\mathbf{x}_3$ and $\mathbf{y}_3$ belong to F_3, we can find two orthogonal transformations Γ_1 , Γ_2 and a non singular linear

transformation L such that $\Gamma_i \mathbf{x}_i = \mathbf{y}_i$, for $i = 1, 2$ and $L\mathbf{x}_3 = \mathbf{y}_3$. Therefore G is generated by two subgroups G_1 and G_2 with $G_1 = \{g = \sigma : \quad \sigma \in R^+\}$, and

$G_2 = \{g = \Gamma_1 \otimes \Gamma_2 \otimes L : \ \Gamma_i \in O_n, \text{with } \Gamma_i F_i \subset F_i, i = 1, 2; L F_3 \subset F_3, L \text{non singular}\}$.
By applying the previous theorem, we have

$$\mathcal{M}_n(P_0) = \{\mathcal{L}(R\Gamma U), \ R \geq 0, \Gamma = \Gamma_1 \otimes \Gamma_2 \otimes L, U \ \text{ has uniform distribution on } S_{n-1}\}.$$

This is exactly the characterization of all distributions for which the Fisher's statistic has the same distribution.

References

[1] Barndorff-Nielsen, O.E.(1988) Parametric Statistical Models and Likelihood. *Lecture Notes in Statist.*,**50**, Springer.

[2] Bourbaki, N.(1971) Eléments de Mathématique : Topologie générale, Chapitre 1-4.
Hermann, Paris, and Addison-Wesley, Reading, Massaschusettes.

[3] Eaton, Moris L.(1989) *Group Invariance Applications in Statistics. Regional Conference Series in Probability and Statistics* , Vol.**1**, 145 pp.

[4] Fang, K.-T. Kotz, S. NG, K.W.(1990) *Symmetric Multivariate and Related Distributions Chpman and Hall* ,London, New York

[5] Jespersen, Niels Ch. B.(1989) *On the structure of transformation models. Ann. Statist.* **17**, 195-208.

[6] Parthasarathy, K.R.(1967) *Probability Measures on Metric Spaces. Academic Press.*

[7] Rûschendorf, L.(1989) *On random translation models.* Statist. & Probab. Letters *North-Holland* **7** 361-367.

[8] Strassen,V. (1965) *The existence of probability measures with given marginals. Ann. Math. Statist.***36**, 423-439

[9] Strasser,H. (1985) *Mathematical Theory of Statistics. De Gruyter, Berlin.*

[10] 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