KC

1

Alternative proof.

Definition 1. *A parameter θ of a probability measure $p \in \mathcal{P}$, where $\mathcal{P}$ is a set of probability measures, is a function $\mathcal{P} \to \mathbf{R}$.*

Definition 2 (Hoeffding)**.** *A parameter θ is called estimable of degree r, if r is the smallest number for which there is a bounded function $h : \mathbf{R}^r \to \mathbf{R}$ such that, for all $p \in \mathcal{P}$, $\theta(p) = Eh(X_1, \cdots, X_r)$, if X_i are iid random variables sampled according to p.*

[1] introduces a concept of testability which is a 'syntactic sugar' on the concept of estimable parameter.

Definition 3 (testability)**.** *Let $\mathcal{P}$ be a set of probability measures and $\mathcal{P}'$ its complement. Let Θ be a family of estimable parameters which are zero on all elements of $\mathcal{P}$. $\mathcal{P}$ is non-testable if and only if Θ contains only 0 function. Otherwise $\mathcal{P}$ is testable.*

Non-testability implies that a statistical test based on U-statistics can not be conducted, which does not rule out, as remarked by G. Peters, possibility of creating tested based on some other estimators e.g. M-estimator. The lemmas we give next are a used in the proof of the Theorem 2, stating non-testability of conditional independence.

- For integrable $f(x, y, z)$, we denote $f(x, z) = \int_R f(x, y, z)dy$

Let $f(x, y, z)$ be a real valued integrable function. If we skip some of the arguments of f and write e.g. $f(x, z)$ we mean function created by integrating the mising variables out, in this particular example $\int_R f(x, y, z)dy$.

- I be a set of all intervals on the real line
- T - family of all functions $R \to I^2$.
- $\mathcal{G} = \left\{ \mathbf{1}\{(x, y) \in t(z)\}p(z)\frac{1}{vol(t(z))} | t \in T, p \in P \right\}$. *where P is set of all densities on R.*
- If $g \in G$, then $g(x, y, z)g(z) = g(x, z)g(y, z)$.

Lemma 1. *Let $h : R^r \to R$, $h \in C^b$, h symmetric. If for all densities p the integral*

$$\int h(z_1, \cdots, z_r) \prod_{i=1}^{n} p(z_i) dz_i = 0, \tag{1}$$

then h is equal to zero.

Proof. Let $p(z) = \sum_{i=1}^{r} \alpha_i p_i(z)$, where $\sum_{i=1}^{r} \alpha_i = 1$, $\alpha_i > 0$. *If we plug in polynomial to the equation 1 and perform integration we obtain a sum o the following form*

$$f(\alpha) = \sum_{1 \leqslant i_1, \cdots, i_r \leqslant r} \alpha_{i_1} \cdots \alpha_{i_r} \beta_{i_1, \ldots, i_r}$$

where α is a vector of α_i and β coefficients are given by

$$\beta_{i_1, \ldots, i_r} = \int h(z_1, \cdots, z_r) \prod_{i=1}^{r} p_{i_r}(z_i) dz_i.$$

$f(\alpha)$ is a homogoneus, polynomial of r variables. Since β is invariant to permutation of indexes $i_1, ..., i_r$, $f(\alpha)$ can be written

$$f(\alpha) = \sum_{1 \leqslant i_1 \leqslant \cdots \leqslant i_r \leqslant r} \alpha_{i_1} \cdots \alpha_{i_r} C(i_1, ..., i_r) \beta_{i_1, \ldots, i_r}$$

Where $C(i_1, ..., i_r)$ is number of permutations of the set $i_1, ..., i_r$ (e.g. for (1,1,...,1,2) its r). By the assumptions, we know that

$$f(\alpha) = 0$$

on r-dimensional simplex.

This imples that f must be a zero poynomial; otherwise the polynomial

$$b(\alpha) = 1 - \sum_{i=1}^{r} \alpha_i = 1$$

would be an divisor of w, which is impossible since w is homogoneus .

$$\beta_{1, \cdot, r} = 0$$

$$\int h(z_1, \cdots, z_r) \prod_{i=1}^{r} p_i(z_i) dz_i = 0.$$

Since p_i are arbitrary and h is continous we conclude that h is zero. □

Theorem 1. *If a continous, bounded symmetric functions $h : R^{3n} \to R$ is such that*

$$\int_{R^{3n}} h(x_1, y_1, z_1, \cdots, x_n, y_n, z_n) \prod_{i=1}^{n} p(x_i, y_i, z_i) dx_i dy_i dz_i = 0$$

for all $p \in \mathcal{G}$, then $h = 0$ everywhere.

Proof. Assume to the contrary that $h \neq 0$ exists. *By the Fubini theorem*

$$\int \left(\int h(x_1, y_1, z_1, \cdots, x_r, y_r, z_r) \prod_{i=1}^{r} 1\{(x_i, y_i) \in t(z_i)\} \right) \prod_{i=1}^{r} p(z_i) \frac{1}{vol(t(z_i))} \quad (2)$$

Inner integral depends only on t and does not depend on $p(z)$. Denote

$$g_t(z_1, \cdots, z_r) = \int h(x_1, y_1, z_1, \cdots, x_r, y_r, z_r) \prod_{i=1}^{r} 1\{(x_i, y_i) \in t(z_i)\} dx_i dy_i$$

By Lemma 1 $g_t = 0$. *It is clear that g_t is symmetric with respect to its arguments and since $p(z)$ can be chosen arbitrarily, assumptions are met.* For all t

$$\int h(x_1, y_1, z_1, \cdots, x_r, y_r, z_r) \prod_{i=1}^{r} 1\{(x_i, y_i) \in t(z_i)\} dx_i dy_i = 0 \quad (3)$$

Denote $(a_i, b_i) \times (c_i, d_i) := t(z_i)$, so

$$\int_{a_1}^{b_1} \int_{c_1}^{d_2} \cdots \int_{a_r}^{b_r} \int_{c_r}^{d_r} h(x_1, y_1, z_1, \cdots, x_r, y_r, z_r) \prod_{i=1}^{r} dx_i dy_i = 0. \quad (4)$$

Since h is continuous $h = 0$. □

Notation

- $\mathbf{S} = \{(X_i, Y_i, Z_i)\}_{1 \leqslant i \leqslant n}$, iid.
- $\mathbf{S} \sim p$ means $(X_i, Y_i, Z_i) \sim p$.
- Null hypotesis and alternative hypotesis is a set of probability measures.
- statistical test: $\psi : R^{3n} \to 0, 1$
- ψ accepts if $\psi(\mathbf{D}) = 0$, otherwise rejects.
- $\Lambda(\mathcal{A}) = \sup_{\mathbf{S} \sim p \in \mathcal{A}} p(\psi(\mathbf{X}) = 1)$
- Type one $\Lambda(\mathcal{H}_0)$

Statement 1. *If ψ controls type one error on level alpha, then there exist a bounded, symmetric function $h : R^{3n} \to R$ such that for all $p \in \mathcal{P}$*

$$Eh(S) = \int_{R^{3n}} h(s) dp(s) \leqslant 0$$

Proof. Put $h(s) = \psi(s) - \alpha$. Since $\psi(s) = 1\psi(s) = 1$,

$$Eh(S) = E1\psi(s) = 1 - \alpha = p(\psi(s) = 1) - \alpha \leqslant 0.$$

h is symmetric since ψ is symmetric. □

Statement 2. *Suppose $U = \psi = 1$ is open and has finite Lensesgue measure. If $\sup_{p \in P'} \|p\|_\infty = C \leqslant \infty$, there exists continous, bounded h, such that for all $S \sim p \in P'$*

$$Eh(S) \leqslant \alpha.$$

Proof. Let F be a complement of F and define $f_n = min(1, nd(s, F))$ (draw). $f_n \in C_b(R^{3n})$, is a pointwise convergent (to $1U$), non-decreasing sequence od positive functions, so by Lebesgue's monotone convergence theorem there exists N and small epsilon,

$$|\int (f_N(s) - 1U) ds| \leqslant \frac{\epsilon}{C},$$

and so

$$\sup_{p \in P'} |\int (f_N(s) - 1) p(s) ds| \leqslant \epsilon$$

□

Theorem 2. *If a continous, bounded symmetric functions $h : R^{3n} \to R$ is such that*

$$\int_{R^{3n}} h(x_1, y_1, z_1, \cdots, x_n, y_n, z_n) \prod_{i=1}^{n} p(x_i, y_i, z_i) dx_i dy_i dz_i \leqslant 0$$

for all $p \in \mathcal{G}$,

$$\int_{R^{3n}} h(x_1, y_1, z_1, \cdots, x_n, y_n, z_n) \prod_{i=1}^{n} p(x_i, y_i, z_i) dx_i dy_i dz_i \leqslant 0$$

for $\sup_{p \in P'} \|p\|_\infty = C \leqslant \infty$.

References

[1] Wicher P Bergsma. *Testing conditional independence for continuous random variables.* Eurandom, 2004.