NOTES ON TWO NOTIONS OF PARAMETRIC MODELS

H.R.LAW

1. Introduction

This note basically clarifies two notions: Identifiability and estimability. Actually, identifiability is a notion about parametric families, while the estimability is a notion about parametric functions.

2. Identifiable Parametric Family

Definition 1. (Identifiable parametric family[Shao] pp. 93) A parametric model refers to the assumption that the population P is in a given parametric family. A parametric family $\{P_\theta : \theta \in \Theta\}$, which is a collection of probability measures, is said to be identifiable if and only if $\theta_1 \neq \theta_2, \theta_i \in \Theta \Longrightarrow P_{\theta_1} \neq P_{\theta_2}$.

Corollary 2. *A parametric family* $\{P_\theta : \theta \in \Theta\}$ *is identifiable if and only if the mapping* $\Theta \to \{P_\theta : \theta \in \Theta\}, \theta \mapsto P_\theta$ *is injective.*

The estimability is often talked about in the background of linear models. Linear models of the form:

$$\begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \cdots \\ \epsilon_n \end{pmatrix}, \epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$$

This model does assume an underlying parametric model

$$\left\{ P_{(\mu_1, \cdots, \mu_p)} = \left(P_{Y_1}, \cdots, P_{Y_p} \right) : (\mu_1, \cdots, \mu_p) \in \Lambda \right\}$$

where the $Y_i \sim P_{Y_i}$ are the observed response variables. The classical Gaussian assumption is $Y_i \overset{i.i.d}{\sim} N\left(\mu_i, \sigma^2\right) \Longleftrightarrow P_{Y_i} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_i - \mu_i)^2}{2\sigma^2}}$. Therefore we can talk about the identifiability for the parametric model

$$\left\{ P_{(\mu_1, \cdots, \mu_p)} = \left(P_{Y_1}, \cdots, P_{Y_p} \right) : (\mu_1, \cdots, \mu_p) \in \Lambda \right\}$$

. Or we can also talked about their joint distribution, which is a multivariate normal distribution. Due to the factorization theorem of independent uni-variate variables, it is just equivalent to discussing about their own distributions separately.

Proposition 3. *The parametric model* $\left\{ P_{(\mu_1, \cdots, \mu_p)} = \left(P_{Y_1}, \cdots, P_{Y_p} \right) : (\mu_1, \cdots, \mu_p) \in \Lambda \right\}$ *is identifiable.*

Proof. If $(\mu_1, \cdots, \mu_p) \neq \left(\mu_1', \cdots, \mu_p'\right)$ then there exists at least one $\mu_k \neq \mu_k'$. And we can examine that $P_{Y_k} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_k - \mu_k)^2}{2\sigma^2}} \neq \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_k - \mu_{k'})^2}{2\sigma^2}}$ with two different parameters. And refer to the definition, we are done. $\square$

3. Estimable Parameter Functions

Definition 4. (Estimable parametric functions [Scheffe] pp.13) A parametric function, which is a *linear* function uses parameters as variables, is defined to be a linear function of the unknown parameters $\{\mu_1, \cdots, \mu_p\}$ with known constant coefficients $\{c_1, \cdots, c_p\}$. An usual form is like $\psi = \sum_{k=1}^{p} c_k \mu_k$. And it is called an estimable function if it has an unbiased linear estimate. [1]

This is the usual definition of estimable functions we can find for linear model's parameters. However, J.Shao provided a more general definition for estimability.

Definition 5. (Estimable parametric functions [Shao] pp.161) If there exists an *unbiased* estimator of the parameter θ of parametric family $\{P_\theta : \theta \in \Theta\}$, then θ is called an estimable parameter.

The discussion in [Shao] pp.183 told us that in the linear model:

$$\begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \cdots \\ \epsilon_n \end{pmatrix}$$

And by the generalized inverse version of solutions to the normal equations gives us the LSE of the coefficients:

$$\begin{pmatrix} \hat{\beta}_1 \\ \cdots \\ \hat{\beta}_p \end{pmatrix} = \left(\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix}^{\tau} \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} \right)^{-} \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix}^{\tau} \begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix}$$

Proposition 6. *If the design matrix* $\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix}$ *is of rank* $r < p$*, then* $\left\{ P_{(\mu_1, \cdots, \mu_p)} = \left(P_{Y_1}, \cdots, P_{Y_p} \right) : (\mu_1, \cdots, \mu_p) \in \Lambda \right\}$ *is NOT identifiable.*

Proof. It suffices to show a counter-example. Now the distribution, with parameter plugged in, of Y_i is $N\left(\sum_{j=1}^{p} X_{ji}\beta_j = \mu_i, \sigma^2 \right)$.[2] And correspondingly, $P_{Y_i} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(Y_i - \mu_i)^2}{2\sigma^2}}$. However, since the design matrix is not of full rank, we know from linear algebra that there exists $\begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} \neq \begin{pmatrix} \beta_1' \\ \cdots \\ \beta_p' \end{pmatrix}$ while $\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} \begin{pmatrix} \beta_1' \\ \cdots \\ \beta_p' \end{pmatrix}$. [3]. Now we know $\mu_i = \sum_{j=1}^{p} X_{ji}\beta_j = \sum_{j=1}^{p} X_{ji}\beta_j' = \mu_i' \Longrightarrow P_{(\beta_1, \cdots, \beta_i, \cdots, \beta_p)} = P_{(\mu_1, \cdots, \mu_i, \cdots, \mu_p)} = P_{\left(\mu_1, \cdots, \mu_i', \cdots, \mu_p\right)} = P_{\left(\beta_1, \cdots, \beta_i', \cdots, \beta_p\right)}$. In other words, there are "redundant" parameters. □

---

[1] Actually, Scheffé referred the reader to a very rare article *The fundamental theorem of linear estimation*(1944) by R.C.Boses, which I can no longer found.

[2] If we plug in these parameter LSE, then it should be $Y_i \sim N\left(\sum_{j=1}^{p} X_{ji}\hat{\beta}_j = \hat{\mu_i}, \sigma^2 \right)$

[3] Say, $\begin{pmatrix} \beta_1' \\ \cdots \\ \beta_p' \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \cdots \\ \gamma_p \end{pmatrix}, \begin{pmatrix} \gamma_1 \\ \cdots \\ \gamma_p \end{pmatrix} \in Null \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} \neq \emptyset$ will play the role.

So, we must find $\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} = \mathbb{X}_{\mathrm{n}\times\mathrm{r}} Q_{r\times p}$ where we throw out those linear dependent observations to yield $\mathbb{X}$. Due to this QR-decomposition, we can see $Q\begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix}$ is identifiable.[4]

If we try to estimate a non-identifiable parameter, then the result is meaningless. But what if we try to estimate a non-estimable parameter? First we try to prove:

Theorem 7. *([Shao] pp.184 Theorem 3.6.(iii)) If* $l \notin$ *Row* $\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix}$*,then* $l^\tau \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix}$ *is NOT estimable.*

Proof. It suffices to prove there is no unbiased estimator of the parameter function $l^\tau \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix}$. Assume the contrary, there is some $h(\begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix}, \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix})$ such that $E\left(h(\begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix}, \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix})\right) = l^\tau \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix}$. We can calculate the expectation directly by:

$$\int h(\begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix}, \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix}) d\begin{pmatrix} P_{Y_1} \\ \cdots \\ P_{Y_n} \end{pmatrix} = l^\tau \begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix}$$

Differentiate with respect to the vector $\begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix}$

$$\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix}^\tau \int h(\begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix}, \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix})(\begin{pmatrix} y_1 \\ \cdots \\ y_n \end{pmatrix} - \begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix}\begin{pmatrix} \beta_1 \\ \cdots \\ \beta_p \end{pmatrix})d$$

□

Corollary 8. *The parameter function is estimable if and only if it is of the form* $\psi = \sum_{k=1}^{p} c_k \mu_k$.

Proof. The "only if" part is directly from the contra-positive of the previous theorem since $\mu_k = \sum_{j=1}^{p} X_{jk}\beta_j$, $\psi = \sum_{k=1}^{p} c_k(\sum_{j=1}^{p} X_{jk}\beta_j)$. The "if" part can be directly verified since $E(\sum_{j=1}^{p} X_{ji}\beta_j) = \mu_i$ due to the Gaussian assumption and the expectation operator is linear over real numbers. □

Therefore, as long as we are concerned with only linear functions of parameters (i.e. parameter functions), then those only estimable functions are exactly those

[4][Shao] pp.183:"In many applications, we are interested in estimating some linear functions of β, i.e., $\theta = l^\tau\beta$ for some $l \in \mathcal{R}^p$. From the previous discussion, however, estimation of $l^\tau\beta$ is meaningless unless $l = Q^\tau c$ for some $c \in R^r$ so that $l^\tau\beta = c^\tau Q\beta = c^\tau\tilde{\beta}$"

linear combinations of group means μ_i. Although a parameter function of this form is necessarily estimable, it does not mean that this parameter function is identifiable since more than one groups can be "redundant". Thus we have the following closing conclusion:

Theorem 9. *In linear model with Gaussian assumptions, for a linear function of parameters τ_i, rank* $\begin{pmatrix} X_{11} & \cdots & X_{p1} \\ \cdots & & \cdots \\ X_{1n} & \cdots & X_{pn} \end{pmatrix} = r \leq p$

(i) A parameter function ψ is estimable iff $\psi = \sum_{k=1}^{p} c_k \mu_k$

(ii) The parametric family P_μ is identifiable iff μ is a parameter of rank r (i.e.$rank(\mu\mu^\tau) = r$).

So it is safe to say that a parameter function is estimable in a linear model iff it is a linear combination of those group means; the whole parametric family is identifiable iff the parameter is "non-redundant".

Try to estimate a non-estimable parameter function is always inaccurate in sense of biasedness.

Try to estimate the parameters a non-identifiable parametric family is always meaningless for telling an exact distribution from this family.

References

[Casella&Berger] G.Casella and O.Berger, Statistical Inference, 2ed, John Wiley&Sons, 2003

[Scheffe] H.Scheffé, The Analysis of Variance, John Wiley&Sons, 1959

[Shao] Mathematical Statistics, 2ed, Springer, 2003